Chapter 1:

1. Fill in the black

1.1. The process of using sample statistics to draw conclusions about true population parameters is called…

statistical inference

Ex:

+ Based on sample results, we estimate that p, the proportion of all U.S. adults who are in favor of stricter gun control, is 0.6.

+ Based on sample results, we are 95% confident that p, the proportion of all U.S. adults who are in favor of stricter gun control, is between 0.57 and 0.63.

1.2. A summary measure that is computed to describe a characteristic from only a **sample of the population** is called…

a statistic

Ex:

+ A **sample** of 120 employees of a company is selected, and the average age is found to be 37 years.

+ **Surveyed 100** students in a university and found that 18 students like to watch movies on weekends.

Ex: **a parameter:**

+ Of **all employees** of the company is selected, and the average age is found to be 37 years

+ **Surveyed all** students in a university and found that 18 students like to watch movies on weekends.

1.3. Which of the following is not an element of descriptive statistical problems?

a) an inference made about the population based on the sample
b) the population or sample of interest
c) tables, graphs, or numerical summary tools
d) identification of patterns in the data

Ex about descriptive statistical:

+ **Surveyed 100** students in a university and found that 18 students like to watch movies on weekends.

+ A **sample** of 120 employees of a company is selected, and the average age is found to be 37 years.

1.4 Which of the following is NOT a reason for the need for sampling?

a) It is usually too costly to study the whole population.
b) It is usually too time consuming to look at the whole population.
c) It is sometimes destructive to observe the entire population.
d) It is always more informative to investigate a sample than the entire population.

1.5 Which of the 4 methods of **data collection** is involved when a person records the use of the Los Angeles freeway system?

a) published sources
b) experimentation
c) surveying
d) observation

Note: An **observation** is a **data collection method**.

1.6 The most frequently occurring data value is …

mode

Ex:

Mark	6	7	8	9
freq	4	8	5	8

Mode: 7 and 9

Chapter 2:

1. Choose randomly 2 balls in the bag that contains 20 balls. How many ways are there to solve problems of sampling without replacement.

There are $C_{20}^{2} = \frac{20 \times 19}{1 \times 2} = 190$ ways.

2. The human sex ratio is the number of males for each female in a population. If the sex ratio in the company is 3, choose randomly four people. Find the probability that they are all males.

Let p be the proportion of males in the company. We have p = ¾ and pro = $(¾)^4$ = 0.3164

3. A player rolls a standard pair of dice. If the sum of the numbers is a 6, the player wins $6. If the sum of the numbers is anything else, the player has to pay $1. What is the expected value for this game?

Outcome	-1	6
Probability	5/6	1/6

Expected = -1*5/6 + 6*1/6 = 1/6 ($)

4. A student buys 3000 integrated circuits (ICs) from supplier A, 5000 ICs from supplier B and 2000 ICs from supplier C. He tested the ICs and found that the conditional probability of an IC being defective depends on the supplier from it was bought. Specifically, given that an IC came supplier A, the probability that it is defective is 0.15; given that an IC came supplier B, the probability that it is defective is 0.2; given that an IC came supplier C, the probability that it is defective is 0.25. If ICs from the three supplier are mixed together and one is selected at random, what is the probability that it is defective?

3000/10 000 x 0.15 + 5000/10 000 x 0.2 + 2000/10 000 x 0.25 = ...

5. Disks of polycarbonate plastic from a supplier are analyzed for scratch and shock resistance. The results from 100 disks are summarized as follows:

	Shock Resistance		
Scratch Resistance		High	Low
	High	50	11
	Low	80	15

Let A denote the event that a disk has no high shock resistance, and let B denote the event that a disk has high scratch resistance. If a disk is selected at random, determine the probability P(B|A)

P(B|A) = 11/(11+15)

6. Let the random variable X be a Poisson distribution with mean of 4.21. Find the probability that X = 4.

pro = $e^{-4.21}$x4.21^4/4!

7. In one town, 70% of adults have health insurance. What is the probability that 6 adults selected at random from the town all have health insurance?

pro = 0.7^6

8. Suppose X has a **hypergeometric** distribution with N=20, n=4, and K=5. Find P(X=3) and P(X ≤ 3).

$$X \sim H(20,5,4)$$

$$P(X=3)=\frac{C_5^3 C_{15}^1}{C_{20}^4};P(X\le 3)=P(X=3)+...+P(X=0)=1-P(X=4)=1-\frac{C_5^4 C_{15}^0}{C_{20}^4}$$

9. Suppose the random variable X has a **geometric** distribution with a mean of 1.6. Find P(X=5)

$$E(X)=\frac{1}{p}\to p=\frac{1}{E(X)}=\frac{1}{1.6}$$

$$P(X=5)=q^4 p=\left(1-\frac{1}{1.6}\right)^4\times\frac{1}{1.6}$$

10. Find the mean and variance of the given probability distribution.

x	P(x)
0	0.1
1	0.2
2	0.1
3	0.3

4	?

$P(4) = 1 - (0.1 + 0.2 + 0.1 + 0.3) = 0.3.$

FX570---> E(X) and Var(X)

11. If X is discrete uniform distributed over the interval [50,100]. Compute the probability that $20 < X < 80$.

$20<X<80$ ---> $50 \le X \le 79$

p=30/51

12. The probability that a radish seed will germinate is 0.1. A gardener plants seeds in batches of 1120. Find the standard deviation for the number of seeds germinating in each batch.

$X \sim B(1120, 0.1)$

Var(X) = npq = 1120x0.1x0.9 --->sd = sqrt(Var(X)) = sqrt(1120x0.1x0.9)=...

Chapter 4:

1. Let X be a continuous uniform distribution over the interval [1,10]. Find the mean, variance of X and P(3<X<20)

$$E(X)=\frac{a+b}{2}=\frac{1+10}{2};\mathrm{var}(X)=\frac{(b-a)^2}{12}=\frac{(10-1)^2}{12}$$

$$P(3<X<20)=P(3<X<10)=\frac{10-3}{10-1}$$

2. If X is a continuous random variable with probability density function f(x) = cx^2 , 0 < x <2 . Find the mean, the variance of X and P(1<X<3)

$$\int_0^2 cx^2dx=1\to c=\frac{1}{\int_0^2 x^2dx}=\frac{3}{8}$$

$$E(X)=\int_0^2 x\times\frac{3}{8}x^2dx=\int_0^2\frac{3}{8}x^3dx=\frac{3}{2}$$

$$Var(X)=\int_0^2 x^2\times\frac{3}{8}x^2dx-\left(\frac{3}{2}\right)^2=\frac{3}{20}$$

$$P(1<X<3)=P(1<X<2)=\int_1^2\frac{3}{8}x^2dx=\frac{7}{8}$$

3. Let the random variable X represent the profit made on a randomly selected day by a certain store. Assume that X is Normal with mean 450 and standard deviation 20. Compute P(X > 460) and P(455>X > 440)

P(X>460) = 1 – F((460-450)/20)=1-F(0.5)

P(455>X > 440) = F((455-450)/20) – F((440-450)/20)=...

Calculate F, we use casio or F is given

4. The continuous random variable X has probability density function is f(x)=e^{-x}, x>0. Find P(X=2010)

P(X=2010) = 0

5. Assume X is normally distributed with a mean of 7 and standard deviation of 2. Find: P(X<9) and P(5<X<9)

$$X \sim N\left(7, 2^2\right)$$

$$P\left(a < X < b\right) = F\left(\frac{b-\mu}{\sigma}\right) - F\left(\frac{a-\mu}{\sigma}\right)$$

$$P\left(X < 9\right) == F\left(\frac{9-7}{2}\right) = F\left(1\right) = P\left(Z < 1\right)$$

$$P\left(5 < X < 9\right) = F\left(\frac{9-7}{2}\right) - F\left(\frac{5-7}{2}\right)$$

Chapter 6-7

1. Ten measurements were made on the inside diameter of forged piston rings used in an automobile engine. The data (in millimeters) are 7; 7.4; 7.5; 7.2; 7.8; 7.6; 7.4;7.1; 7.9 and 7.3. Calculate the mean and sample standard deviation

casio

2. The data are 7; 7.4; 7.5; 7.2; 7.8; 7.6; 7.4;7.1; 7.9 and 7.3. Find mode, IQR and outliers

```
> summary(c(7,7.4,7.5,7.2,7.8,7.6,7.4,7.1,7.9,7.3))
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  7.000   7.225   7.400   7.420   7.575   7.900
```

$$mode: \ 7.4$$

$$IQR = Q_3 - Q_1 = 7.575 - 7.225 = 0.35$$

$$Q_1 - 1.5IQR = 6.7$$

$$Q_3 + 1.5IQR = 8.1$$

$$outliers: x < Q_1 - 1.5IQR \, or \, x > Q_3 + 1.5IQR$$

$$No \, outlier$$

3. Suppose that sample of size n = 15 are selected at random a normal population with mean 80 and standard deviation 30. What is the probability that the sample mean falls in the interval from 75 to 84?

$$SE\ mean = \sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{30}{\sqrt{15}} = 7.75$$

$$\mu_{\overline{X}} = \mu = 80$$

$$P\left(75 < \overline{X} < 84\right) = F\left(\frac{84-80}{7.75}\right) - F\left(\frac{75-80}{7.75}\right)$$

4. Suppose that a random variable X has a continuous uniform distribution with f(x)=0.1 , where x is in [0,10]. Find the mean and variance of the sample mean of a random sample of size n = 40.

$$\mu = \frac{a+b}{2} = \frac{0+10}{2} \qquad Var\left(\overline{X}\right) = \frac{\sigma^2}{n} = \frac{\frac{(b-a)^2}{12}}{40}$$

5. A normal population has mean 1200 and variance 150. How large must the random sample be if we want the standard error of the sample mean to be 2.8?

standard error of the sample mean (SE mean) = sd/sqrt(n) = sqrt(150)/sqrt(n) = 2.8 --->n

Chapter 8

1. The diameter of holes for a cable harness is known to have a normal distribution with standard deviation 0.1 inch. A random sample of size 25 yield an average diameter of 12.3 inch. Find a 94% two - side confidence interval on the mean hole diameter.

94% ---> $z_{\alpha/2} = z_{0.03} = 1.885$

e = $z_{\alpha/2}$ x sd/sqrt(n) = 1.885x0.1/sqrt(25)

confidence interval: [12.3-e; 12.3+e]

2. An article describes the effect of delamination on the natural frequency of beams made from composite laminates. Six such delaminated beams were subjected to loads, and the resulting frequencies were as follows (in hertz): 23; 24; 23.5; 22; 24.5 and 23.5. Calculate a 98% two - sided confidence interval on mean natural frequency.

n =6; mean = ...; s = ... (casio)

98% --->$t_{0.01}^{5}$ = ...(given)

e = $t_{0.01}^{5}$ s/sqrt(n) = ...

CI: [mean –e; mean + e]

3. Of 1000 randomly selected cases of lung cancer, 450 resulted in death within 5 years. Calculate a 96% CI on the death rate from lung cancer.

n = 1000; f = 450/1000 = 0.45; 96% ---> $z_{0.02}$ = 2.055 (given)

e = 2.055x sqrt(f(1-f)/n) = ...

CI: [f-e; f+e]

4. An economist is interested in studying the incomes of consumers in a particular region. The population standard deviation is known to be 250 (USD). A random sample of 50 individuals resulted in an average income of 1000 (USD). What total sample size would the economist need to use for a 99.5% confidence interval if the width of the interval should not be more than 30 (USD)?

e = width/2 = 15

e = $z_{0.005/2}$ x sd/sqrt(n) ---> 15 = 2.807 x 250/sqrt(n) --->n

5. The daily intakes of milk (in ounces) for ten randomly selected people were: 123; 329; 331; 263; 108; 268; 274; 297; 190; 299. Find a 95% confidence interval for the population standard deviation σ.

n=10, s = ...; chis(9)$_{0.025}$ = ... chis(9)$_{0.975}$ = ...

CI: [sqrt((n-1)s^2/ chis(9)$_{0.025}$); sqrt((n-1)s^2/ chis(9)$_{0.975}$)]

Note: 1-size

$$z_{\frac{\alpha}{2}} \to z_{\alpha}; t_{\frac{\alpha}{2}} = t_{\alpha}$$

$$: \ Upper = \bar{x} + z_{\alpha}\frac{\sigma}{\sqrt{n}}; f + z_{\alpha}\frac{\sqrt{f(1-f)}}{\sqrt{n}}$$

$$Lower = \bar{x} - z_{\alpha}\frac{\sigma}{\sqrt{n}}; f - z_{\alpha}\frac{\sqrt{f(1-f)}}{\sqrt{n}}$$

Chapter 9

1. An engineer who is studying the tensile strength of a steel alloy intended for use in golf club shafts knows that tensile strength is approximately normally distributed with standard deviation 10 psi. A random sample of 25 specimens has a mean tensile strength of 39 psi. Test the hypothesis that mean strength is 45 psi. Use α=0.01.

n =25, mean = 39, sd = 10; mu_0 = 45

z = (mean – mu_0)*sqrt(n)/sd = (39 – 45)*5/10 = -3

$z_{\alpha/2}$ = 2.575

|z| > $z_{\alpha/2}$: reject the null hypothesis

Note: $P-value = 2F\left(-2\right)$

2. A random sample of 350 circuits generated 30 defectives. Use the data to test H0:p=0.1 versus H1:p>0.1

n = 350; f = 30/350

z =(f – p)sqrt(n)/sqrt(p(1-p))

p_{value} = 1- F(z)

3. The mean water temperature downstream from a power plant cooling tower discharge pipe should be no more than 100°F. Past experience has indicated that the standard deviation of temperature is 2°F. The water temperature is measured on ten randomly chosen days, and the average temperature is found to be 97°F. Test the hypotheses $H_0{:}\mu = 100$, $H_1{:}\mu < 100$ by using p_value

n= 10, mean = 97, sd = 2

z =(mean – μ)*sqrt(n)/sd = ...

p_{value} = F(z)=...

4. What are type I error and type II error?

A type I error is the rejection of a true null hypothesis

A type II error is the failure to reject a false null hypothesis

5. An Izod impact test was perform on 25 specimens of PVC pipe. The sample standard deviation was 2.2 Test H_0: $\sigma = 2$, H_1: $\sigma \neq 2$, using $\alpha = 0.05$.

n =25, s = 2.2

$chis_{test} = (n-1)s^2/ \sigma^2$ =...

chis1 = a; chis2 = b (given)

accept the null hypothesis if $a < chis_{test} < b$

Note:

z<0: p-value 2-size = 2F(z), p-value 1-size = F(z)

z>0: p-value 2-size = 2F(-z), p-value 1-size = F(-z)

Chapter 11

1. Which correlation coefficient represents the weakest association between the X and Y variables?

Select one:

(a) $r = 0.50$

(b) $r = -0.20$

(c) $r = 0.90$

(d) $r = 0.10$

2. A random sample of n=38 observations was made on the time to failure of an electronic component and the temperature in the application environment in which the component was used. Given that r = 0.35 Test the hypothesis that H_0: $\rho = 0$ versus H_1: $\rho \neq 0$ with α=0.05

n =38; r = 0.35

t = r*sqrt(n-2)/sqrt(1-r^2) = ...

given $t^{36}_{0.025}$; reject the null hypothesis if $|t| > t^{36}_{0.025}$

3. The paired data below consist of the test scores of 6 randomly selected students and the number of hours they studied for the test.

Hours	5	10	15	20	10	5
Score	4	8	10	12	9	3

28.1 Find the value of the linear correlation coefficient r

r = ...(casio)

28.2 Find the equation of the regression line.

Y = A + BX (casio)

28.3 Test the hypothesis that H_0: $\rho = 0$ versus H_1: $\rho \neq 0$ with α=0.05

n = 6; r =...; t = r*sqrt(n-2)/sqrt(1-r^2) = ...

given $t^{36}_{0.025}$; reject the null hypothesis if $|t| > t^{36}_{0.025}$

Note: degree freedom: df = n-2