CHAPTER 5

Methods of Collecting Data

NOTE: Both methods involve observations, but...

Observational	An observational study finds the relationship between changes that already exist. • Easier to perform • More difficult to draw cause/effect conclusions from
Experimental	An experiment requires you to observe what happens when you make some sort of change. • Can be unethical • Subjects randomly assigned to groups • Possible to draw cause/effect conclusions

Types of Samples

Simple Random Sample	A sample selected in such a way that every possible sample with the same number of observations is equally likely to be chosen.
Stratified Random Sample	Obtained by separating the population into mutually exclusive sets, or strata, and then drawing simple random samples from each stratum.
Cluster Sampling	A simple random sample of groups or clusters of elements.

Sampling vs Non-Sampling Error

Sampling Error	Difference between what a sample predicts and the true population paramter. Occurs because a sample is being used to make inferences about an entire population.
Non-Sampling Error	Errors resulting from mistakes made when collecting data. • Responses being recorded incorrectly • Some participants not responding • Biased sample

CHAPTER 7

Discrete Random Variables (Ch7) vs. Continuous Random Variables (Ch8)

Random variable	A function or rule that assigns a number to each outcome of an experiment.
Discrete random variable	Values are countable, e.g. number of courses, number of siblings, age (in years).
Continuous random variable	Values are uncountable - typically measured quantities, e.g. height, weight, time, speed

Discrete Probability Distributions

Probability distribution A table, formula, or graph that describes the values of a random variable and the probability associated with these values

Requirements for a Distribution of a Discrete Random Variable...

1. $0 \leq P(x) \leq 1$ (for all *x*)
2. $\Sigma P(x) = 1$ (for all *x*)

Expected Value of X (non-binomial) $E(X) = \mu = \sum_{all\ x} xP(x)$

Variance of X (non-binomial) $V(x) = \sigma^2 = \sum_{all\ x} (x-\mu)^2 P(x)$

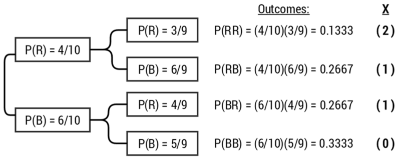
Standard Deviation of X (non-binomial) $\sigma = \sqrt{\sigma^2}$

Example (Non-binomial)

Question:

Two balls are selected randomly without replacement from a jar containing 4 red balls and 6 black balls. Let *X* be the number of red balls selected.

a) Give the probability distribution of x



Outcomes:	X
P(RR) = (4/10)(3/9) = 0.1333	(2)
P(RB) = (4/10)(6/9) = 0.2667	(1)
P(BR) = (6/10)(4/9) = 0.2667	(1)
P(BB) = (6/10)(5/9) = 0.3333	(0)

x	0	1	2
P(x)	**0.3333**	2 x 0.2667 =**0.5334**	**0.1333**

b) Find the expected value and variance of x

x	0	1	2
P(x)	0.3333	0.5334	0.1333
xP(x)	(0)(0.3333) =0	(1)(0.5334) =0.5334	(2)(0.1333) =0.2666
(x-μ)	-0.8	0.2	1.2
$(x-\mu)^2$	0.64	0.04	1.44
$(x-\mu)^2P(x)$	0.2133	0.0213	0.1920

E(x)=μ =ΣxP(x) =0.8

$V(x)=\sigma^2$ $=(x-\mu)^2P(x)$ = 0.4266

Laws of Expected Value	Laws of Variance
1. *E*(*c*) = *c*	1. *V*(*c*) = *0*
2. *E*(*X* + *c*) = *E*(*X*) + *c*	2. *V*(*X* + *c*) = *V*(*X*)
3. *E*(*cX*) = *cE*(*X*)	3. *V*(*cX*) = c^2*V*(*X*)

Binomial Distribution

Is it a Binomial Experiment?

1. Fixed number of trials / observations (*n*).
2. Each trial has two possible outcomes: *success and failure*.
3. P(success) = *p*, P(failure) = 1 − *p*.
4. The trials are independent of each other.

Binomial Probability of (x) Number of Successes:

$$P(X = x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$$

Mean or Expected X (Binomial)	Variance of X (Binomial)	Standard Deviation of X (Binomial)
$\mu = np$	$\sigma^2 = np(1-p)$	$\sigma = \sqrt{np(1-p)}$

Using the Binomial Table

Use the binomial tables whenever you can!

When can you?

1. The table has the same number of trials / observations (*n*) as in the question.
2. The table has a column with the exact probability of succes (*p*) as in the question.

P(x) is <u>NOT</u> listed on the table!... P(k) is

What you're asked for:	What to find on the table:
$P(X < x) =$	$P(k = [x-1])$
$P(X \leq x) =$	$P(k = [x])$
$P(X = x) =$	$P(k = [x]) - P(k = [x-1])$
$P(X \geq x) =$	$1 - P(k = [x-1])$
$P(X > x) =$	$1 - P(k = [x])$
$P(x_1 < X < x_2) =$	$P(k = [x_2 - 1]) - P(k = [x_1])$
$P(x_1 \leq X < x_2) =$	$P(k = [x_2 - 1]) - P(k = [x_1 - 1])$
$P(x_1 < X \leq x_2) =$	$P(k = [x_2]) - P(k = [x_1])$
$P(x_1 \leq X \leq x_2) =$	$P(k = [x_2]) - P(k = [x_1 - 1])$

CHAPTER 10

Definitions

Point Estimator:	A ***single value*** that estimates an unknown *population parameter.*
Interval Estimator:	A ***range of values*** that estimate and unknown *population parameter.*

Qualities of a Good Estimator

Unbiased:	The average (expected) value of the estimator equals the population parameter being estimated.
Consistency:	The estimator gets closer to the population parameter as the sample size increases.
Relative Efficiency:	If two unbiased estimators are available as estimators, choose the one with less variability.

Interpreting a Confidence Interval

"The mean is estimated to fall between (*LCL) and (*UCL**). This type of estimation is correct (*confidence level**)% of the time."**

**Fill in the values*

Confidence Interval Estimator of μ

Lower Confidence Limit (LCL):	$\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$
Upper Confidence Limit (UCL):	$\bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$
Sample Size (given the bound on error of estimation B) :	$n = \left(\frac{z_{\alpha/2}\sigma}{B}\right)^2$

Common Confidence Levels and $z_{\alpha/2}$

$1-\alpha$	α	$\alpha/2$	$z_{\alpha/2}$
.90	.10	.05	$z_{.05} = 1.645$
.95	.05	.025	$z_{.025} = 1.96$
.98	.02	.01	$z_{.01} = 2.33$
.99	.01	.005	$z_{.005} = 2.575$

Factors Affecting WIDTH of Confidence Intervals

As sample size *n* ↑...	width ↓ (GOOD!!!)
As sample size *n* ↓...	width ↑ (wide intervals are BAD)
As confidence level (1-α) ↑...	width ↑ (wide intervals are BAD)
As confidence level (1-α) ↓...	width ↓ (loss of confidence is BAD)
As variance ↑...	width ↑ (BAD: can't adjust variance)
As variance ↓...	width ↓ (BAD: can't adjust variance)
As sample mean ↑...	width is unaffected
As sample mean ↓...	width is unaffected

STATS
DOESN'T SUCK

CHAPTER 8

Probability Density Functions

Continuous Random Variable

- uncountable (infinite) number of values
- the probability of each individual value is virtually 0

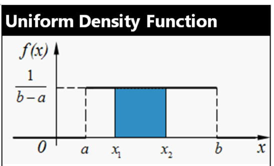
Probability Density Function

- gives probabilities the variable takes on different values

Requirements: (range is *a* ≤ *x* ≤ *b*)

1. *f*(*x*) ≥ 0 for all *x* between *a* and *b*
2. The total area under the curve between *a* and *b* is 1.0

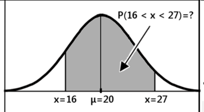
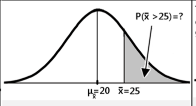
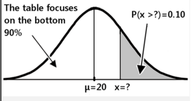
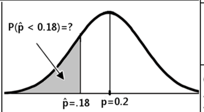
Uniform Density Function



$P(x_1 < X < x_2) =$

$= \text{Base} \times \text{Height}$

$= (x_1 - x_2) \times \frac{1}{b-a}$

<table>
<tr><th colspan="2">Normal Distribution Probabilities</th><th>Ch.8 (x) example</th><th>Ch.9 ($\bar{x}$) example</th><th>REVERSE Ch.8 (Given a probabilty and asked to find boundary)</th><th>Ch.9 ($\hat{p}$) example</th></tr>
<tr><td>QUESTION TYPE</td><td>Given one or two boundaries, find the probability that x (or $\bar{x}$)is less than, greater than, or between them.</td><td>Given: μ = 20, σ = 4 Find the probability that x is between 16 and 27?</td><td>Given: μ = 20, σ = 24, n = 36. Find the probability that $\bar{x}$ is greater than 25?</td><td>μ = 20, σ = 4 What value of x separates the top 10%?</td><td>Probability of success in a binomial experiment is 20%. Find the probability that the proportion of success in a sample of 400 is less than 18%</td></tr>
<tr><td>STEP 1</td><td>Draw the normal distribution.
• Label the mean μ
• Label (each) x (or $\bar{x}$)
• Shade area of interest</td><td>
</td><td>
</td><td>
</td><td>
</td></tr>
<tr><td>STEP 2</td><td>Convert (each) x (or $\bar{x}$) to a z-score as follows:
(Ch.8) $z = \frac{x-\mu}{\sigma}$
(Ch.9) $z = \frac{\bar{x}-\mu}{\sigma_{\bar{x}}}$ $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$</td><td>$z = \frac{16-20}{4} = \frac{-4}{4} = -1$
$z = \frac{27-20}{4} = \frac{7}{4} = 1.75$</td><td>$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{24}{\sqrt{36}} = 4$
$z = \frac{\bar{x}-\mu}{\sigma_{\bar{x}}}$
$z = \frac{25-20}{4} = 1.25$</td><td>**REVERSE STEPS 2+3**
The table requires we use the 'less than'area so...
1-.10=0.90
<table>
<tr><th>Z</th><th>.00</th><th>.01</th><th>...</th><th>.08</th><th>.09</th></tr>
<tr><td>0.0</td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>0.1</td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>...</td><td></td><td></td><td></td><td>(closest)↓</td><td></td></tr>
<tr><td>1.2</td><td></td><td></td><td></td><td>.8997</td><td></td></tr>
</table></td><td>$z = \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}}$
$z = \frac{0.18-0.20}{\sqrt{\frac{0.20(1-0.20)}{400}}} = -1$</td></tr>
<tr><td>STEP 3</td><td>Look up (each) area on the correct (- ,+) z-table:</td><td>
<table>
<tr><th>Z</th><th>.00</th><th>.01</th></tr>
<tr><td>-3.0</td><td></td><td></td></tr>
<tr><td>...</td><td></td><td></td></tr>
<tr><td>-1.1</td><td></td><td></td></tr>
<tr><td>-1.0</td><td>.1587</td><td></td></tr>
</table>
<table>
<tr><th>Z</th><th>...</th><th>.05</th></tr>
<tr><td>0.0</td><td></td><td></td></tr>
<tr><td>...</td><td></td><td></td></tr>
<tr><td>1.7</td><td></td><td>.9599</td></tr>
<tr><td>1.8</td><td></td><td></td></tr>
</table></td><td>
<table>
<tr><th>Z</th><th>.00</th><th>.01</th><th>...</th><th>.05</th><th>.06</th></tr>
<tr><td>0.0</td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>0.1</td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>...</td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>1.2</td><td></td><td></td><td></td><td>.8944</td><td></td></tr>
<tr><td>1.3</td><td></td><td></td><td></td><td></td><td></td></tr>
</table></td><td>Convert z to x, $\bar{x}$, or $\hat{p}$
$x = z\sigma + \mu$ ...Ch. 8
$\bar{x} = z\sigma_{\bar{x}} + \mu$ ...Ch. 9 ($\bar{x}$)
$\hat{p} = z\sigma_{\hat{p}} + p$ ...Ch. 9 ($\hat{p}$)

$x = z\sigma + \mu$
$x = (1.28)(4) + 20$
$x = 25.12$</td><td>
<table>
<tr><th>Z</th><th>.00</th><th>.01</th></tr>
<tr><td>-3.0</td><td></td><td></td></tr>
<tr><td>...</td><td></td><td></td></tr>
<tr><td>-1.1</td><td></td><td></td></tr>
<tr><td>-1.0</td><td>.1587</td><td></td></tr>
</table></td></tr>
<tr><td>STEP 4</td><td>P('less than') = area
P('greater than')=1 - area
P('between')= big area − small area</td><td>P(16 < x < 27) =
= Big area – Small area
= .9599 – .1587
= 0.8012</td><td>$P(\bar{x} > 25)$ = 1 − area
= 1 − 0.8944
= 0.1056</td><td>x=25.12 separates the top 10% from the bottom 90%</td><td>$P(\hat{p} < 0.18)$ = area
$P(\hat{p} < 0.18)$ = 0.1587</td></tr>
</table>

CHAPTER 9

Sampling Distribution of $\bar{X}$

... shows the probabilities for all possible sample means for a given sample size (n).

Mean	Standard Error
$\mu_{\bar{x}} = \mu$	$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ (standard deviation of the sampling distribution)

Central Limit Theorem:

If *X* is normal, then *X* is normal. If *X* is nonnormal, then *X* is approx normal for sufficiently large sample sizes.

Convert $\bar{X}$ to a z-score: $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$

Sampling Distribution of $\bar{X}_1$-$\bar{X}_2$

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1{}^2}{n_1} + \frac{\sigma_2{}^2}{n_2}}}$$

Sampling Distribution of $\hat{p}$

$\hat{p}$ is approx normally distributed provided that *np* and *n*(1 − *p*) are greater than or equal to 5.

Expected $\hat{p}$	Standard Error of $\hat{p}$
$E(\hat{p}) = p$	$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$

Convert $\hat{p}$ to a z-score: $Z = \frac{\hat{p}-p}{\sqrt{p(1-p)/n}}$ $\hat{p} = \frac{X}{n}$