Cumulative AP Practice Test 4
(Free Response)

AP4.41

2 Sample t-test for the difference means ($\mu_1 - \mu_2$)

* t-test since the population σ is unknown
* Use $\alpha = .05$

μ_1 = the true mean difference in electrical potential – diabetic mice
μ_2 = " " " " – normal mice

$H_0: \mu_1 = \mu_2$
$H_A: \mu_1 \neq \mu_2$

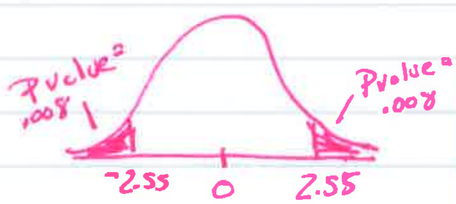
This work will give full credit

USED $\boxed{\text{2-SAMPTTEST}}$

Diabetic mice: $N = 24$
$\bar{x}_1 = 13.09 \quad Sx_1 = 4.839$

Normal Mice: $N = 18$
$\bar{x}_2 = 10.022 \quad Sx_2 = 2.915$

Pooled ALWAYS NO



df (technology) = 38.46
$t = 2.55$

Conditions

Random – diabetic + normal mice were randomly selected

Independent – ① the 2 populations of mice are independent
② Sampling w/o replacement reasonable there are more than 10(24) = 240 diabetic mice and 10(18) = 180 normal mice

Normal – we were told that graphs of both groups reveled No outliers OR STRONG SKEWNESS

Pvalue = $P(t \leq -2.55)$ or $P(t \geq 2.55) = .015$

Conclude: Since the pvalue (.015) is less than $\alpha = .05$, We reject Ho. We have convincing evidence there is a difference in mean electrical potential between their right hip and front feet

AP 4.42

SRS: 9344 white women 65^+ in 4 STATES

Physically fit as teens – 8.5% cognitive impaired

NOT fit – 16.7% cognitive impaired

(A) p_1 = true proportion cognitive impaired – physically fit teens

p_2 = true proportion cognitive impaired – not physically fit.

always define parameters

$H_0: p_1 = p_2$

$H_A: p_1 \neq p_2$

(b) IF CONDITIONS MET → TEST: 2-Sample Z test for the difference of proportions $(p_1 - p_2)$

(C) Results of test: difference significantly different

The results can not be generalized to all women 65+

1. Only white women were included
2. Sample was limited to 4 states

(D) DEFINITION: Two variables are confounded when their effects on the response variable (measure of cognitive decline) cannot be distinguished from one another.

ANSWERS may vary:

* Teens that were physically fit as teens may
  - have also been better students and mentally stimulated
  - have had better diets
  - etc

* Hence we may NOT BE ABLE TO DETERMINE IF PHYSICAL FIT OR THIS OTHER VARIABLE INFLUENCED COGNITIVE DECLINE

AP 4.43

31% IN FAVOR OF "FAT TAX" → 48%
66% OPPOSED THE "FAT TAX" → 49%

TAX $'s GO TO Health Care due to HIGH OBESITY

ⓐ Bias was introduced by associating Health Care funding with obesity. The extra information about obesity influences (bias) respondendent to change minds and "favor the fat tax"

AN UNBIASED Question - focuses solely on the Question of interest "Do you support or oppose a tax on non-diet sugared soda?"

ⓑ The sample was IN NY at various fast-food restaurants. Choosing these locations provide a bias because people that go to fast food restaurants probably prefer sugary sodas.

AN UN biased sample would be a random sample of all residents in NY state.

ⓒ You would want to poll all states, Since the population of each state is different, we would use a "STRATIFIED RANDOM SAMPLE".

To do this method we would need to know the populations of each state.

* then determine their proportions
* apply these pro proportions to your samplesize
* Once you determine the sample sizes of each state, take a SRS of that state

AP 4.44

3 Faculty Members
% arrive first

Given Faculty Members
% Strong Coffee

W = Mr Worcester = 10% ⟵ .30
C = Mr Currier = 50% ⟵ .10
L = Mr Legacy = 40% ⟵ .20
100%

(a) P(Strong Coffee) = .03 + .05 = .08
= $\boxed{.16}$

Easiest method
to do this is
a tree

P(A and B)

W (.10): .3 → S = (.10)(.3) = .03; .7 → NS = (.1)(.7) = .07
C (.50): .1 → S = (.50)(.1) = .05; .9 → NS = (.5)(.9) = .45
L (.40): .2 → S = (.4)(.2) = .08; .8 → NS = (.4)(.8) = .32

1.00

Pencil #'s are NOT NEEDED to answer?

(b) P(Dr Currier Given Strong) = $\frac{P(\text{Currier and Strong})}{P(\text{Strong})} = \frac{.05}{.16}$
= $\boxed{.3125}$

Know Tree + Tables - this makes probability much easier!

Table

	Strong	Not
MRW	.10(.3) = .03	.07
MRC	.5(.1) = .05	.45
MRL	.4(.2) = .08	.32
Total	.16	.84

easy to fill in

Green Sheet
$P(A|B) = \frac{P(A \cap B)}{P(B)}$

Know these Probability Rules

$P(A|B) = P(A \text{ Given } B)$
$P(A \cap B) = P(A \text{ and } B)$
$P(A \cup B) = P(A \text{ or } B)$ ⇒ $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
Independence Rule: $P(A|B) = P(A)$

Green Sheet

Review of the Least Square Regression Line (LSRL)

AP 4.45

RESPONSE VARIABLE (Y) - SEED WEIGHT (mg)
* We want to predict $Y \rightarrow \hat{y}$ ← always put hat on predicted y

EXPLANATORY VARIABLE (x) - SEED COUNT
* We think X may help explain OR INFLUENCE changes in the response variable we are predicting.

Correlation r measures the strength of the linear association between 2 QUANTITATIVE (NUMERIC) variables.

vs r^2 (coefficient of determination) which measures the strength of the linear model. The fraction of the variation account for by the LSRL model

(a) The scatterplot IS NOT LINEAR AND SHOWS A STRONG CURVED PATTERN. THEREFORE A LINEAR MODEL WILL NOT BE APPROPRIATE IN THIS CASE

(b) 2 New models are proposed - Look FOR 2 criterea to determine which model is better
1. Scatterplot should appear linear
2. residual plot should have NO pattern and have scattered residuals

IN THIS EXAMPLE, MODEL B would be better to predict seed weight from seed count.
1. The scatterplot between ln(weight) and ln(seed count) is a moderately strong, negative, linear association.
2. The residual plot has residuals completely scattered with no pattern.

AP 4.45 CONT

(c) Use Model B to predict seed weight if the Seed COUNT is 3,700.

TO ANSWER THIS QUESTION WRITE LSRL model and clearly show work!

Model: $\widehat{\ln(\text{weight})} = 15.491 - 1.5222 \cdot \ln(\text{seed COUNT})$

$X = 3700 \rightarrow \quad \widehat{\ln(\text{weight})} = 15.491 - 1.5222 \cdot \ln(3{,}700)$

$\widehat{\ln(\text{weight})} = 15.491 - 1.5222 \cdot 8.216$

$= 15.491 - 12{,}507$

$\widehat{\ln(\text{weight})} = 2.984$

To undo ln use "e" in the base $\rightarrow$ $e^{\widehat{\ln(\text{weight})}} = e^{2.984}$

To undo log - Use base 10

$\widehat{\text{weight}} = e^{2.984} = 19.77\text{ mg}$

(d) $r^2 = 86.3$ (Never use adjusted r^2)

(in context) About 86.3% of the variability in the predicted ln(seed weight) is accounted for by the LSRL using ln(seed count).

AP 4.46

diameter (correct) $\longrightarrow N(\underset{\mu}{4\text{ IN}}, \underset{\sigma}{.02\text{ IN}})$

Hourly check SRS $n = 25$

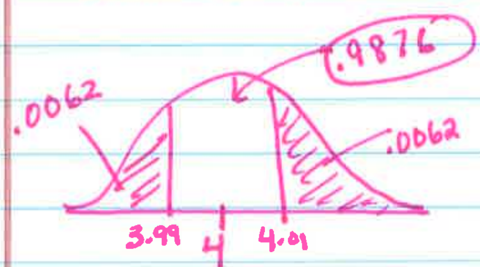
(a) When describing a sampling distribution ALWAYS give: center, spread and shape.

Center $= \mu_{\bar{x}} = \mu = 4\text{ IN}$

Spread $= \sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}} = \dfrac{.02}{\sqrt{25}} = \boxed{.004}$

SHAPE = NORMAL

(b) $P(\bar{x} \leq 3.99 \text{ OR } \bar{x} \geq 4.01) = 1 - .9876 = .0124$



Draw Graph!

Calc Command

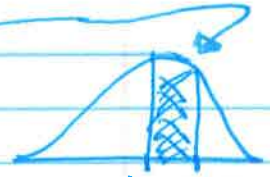
$\text{normalcdf}(\underset{LB}{3.99}, \underset{UB}{4.01}, \underset{\mu_{\bar{x}}}{4}, \underset{\sigma_{\bar{x}}}{.004})$

IF you show calc command make sure to label parameters

AP4.46
CONT

normalcdf
(4, 4.01, 4, .004)

(c) $P(4.00 < \bar{x} < 4.01) = .4938$



(d) DEFINE RV X = the number of samples (OUT OF 5)
where the mean is 4.00 – 4.01.

BINOMIAL Model
Binary Yes
Independent Yes
Number $n = 5$
Success Probability $p = .4938$

X	P(x)
0	.033
1	.162
2	.316
3	.309
4	.150
5	.029

.821 ~~.820~~

.179

1.00

SPECIFIC PROB
binompdf
(5, .4938, 4) = .15
N P X

$P(X \geq 4) = .1798$

Two methods to find

$P(X = 4) + P(X = 5)$
.150 + .029
.179

$P(X \geq 4) =$
$1 - P(X < 4) =$
$1 - P(X \leq 3) =$
$1 - .8201 =$
.1798

Com. PROB
comes from
$-\infty \rightarrow +\infty$

binomcdf
(5, .4938, 3) = .8201
n P X

AP 4.46 (CONT)

(e) OPTION 1: (from b) $P(\bar{x} \le 3.99 \text{ or } \bar{x} \ge .401) = .0124$

OPTION 2: (from d) $P(X \ge 4) = .1798$

OPTION 1 Gives more Convincing evidence that the machine needs to shut down since it has a smaller probability

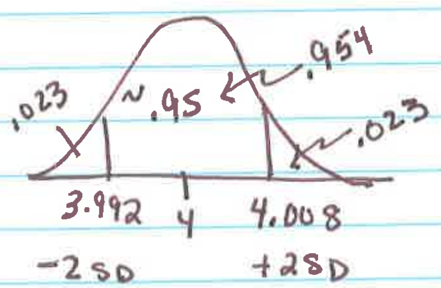
(f) Answers Vary

book: P(5 Consecutive sample means between 4.00 − 4.01) $= (.4938)^5 = .029$

Normalcdf $(4.005, 4.01, 4, .004) = .0994$

book: P(2 Consecutive sample means between 4.005 − 4.01) $=$ $P(4.005 < \bar{x} < 4.01)$ $(.0994)^2 = .01$

USE THE 68-95-99.7 Rule : $\mu_{\bar{x}} = 4$ $\sigma_{\bar{x}} = .004$



Rule: 2 Consecutive sample means that fall out 2 SD's from true mean ($\mu_{\bar{x}} = 4$)

$P(\bar{x} \le 3.992 \text{ or } \bar{x} \ge 4.008) = (.046)^2$

$= .002$

This rule seems like it would be too easy to not catch a problem. Change to ±1 SD

$P(\bar{x} \le 3.996 \text{ or } \bar{x} \ge 4.004)$
$= (.3173)^2 = .101$

This is too high change to 3 samples

FINAL RuLe: $P(\bar{x} \le 3.996 \text{ or } \bar{x} \ge 4.004)^3 = .03$ ← 3 consecutive samples outside 1 SD.