AP CUMULATIVE PRACTICE TEST #3

AP3.1 (E)

$p = .300$

$P(\text{1ST HIT UNTIL 4TH TIME AT BAT}) =$

$P(F) \cdot P(F) \cdot P(F) \cdot P(S) =$

$(.700)(.700)(.700)(.300) =$

$\boxed{(.300)^1 (.700)^3}$

AP3.2 (E)

P(DANDY WINS | GOOD CONDITIONS) = .60

P(GOOD CONDITIONS) = .85

FIND P(Dandy WINS OR GOOD CONDITIONS)

GREEN SHEET: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

TIP: USE TABLE

	GOOD	BAD	
WIN	.60		P(B) DON'T KNOW
LOSE		P(A ∩ B) DON'T KNOW	
	.85		

AP3.3 (D)

MARGIN OF ERROR ± 2% 90% CL

$p = 1/2$ because not given

Green Sheet CI statistic $\pm \underbrace{\text{Critical Value} * \text{SD}}_{ME}$

$$.02 = \overset{1.65}{\cancel{1.645}} \; \hat{z}^* \sqrt{\frac{(1/2)(1/2)}{n}}$$

INV Norm (.05, 0, 1) = 1.645

$$\sqrt{n} = \frac{1.65\,(1/2)}{.02}$$

$$(\sqrt{n})^2 = 41.25^2$$

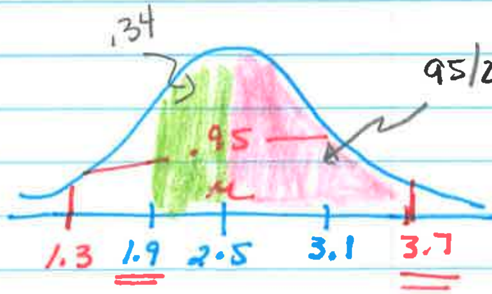
$n = 1701.56$ ROUND UP 1702

AP3.4 (C)

GPA $\sim N(2.5, .6)$

81.5% based on

68 - 95 - 99.7
1SD 2SD 3SD



81.5% = 34% + 47.5%

OR 1.3 to 3.1

AP 3.5 (B)

INCREASE POWER

(1) INCREASE ALPHA (2) INCREASE SAMPLE SIZE

AP 3.6 (D)

VOLUNTARY SAMPLES GIVE VERY LITTLE STATISTICALLY VALID INFO DUE TO BIAS OF THE SAMPLING TECHNIQUE

AP3.7 (C)

	HOLIDAY 1	HOLIDAY 2	
silver	.3	.5	silver
red	.3	.5	blue
pink	.4		
	1.0	1.0	
n =	40	n = 40	

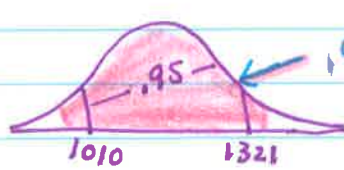
Green Sheet
$\sigma_x = \sqrt{np(1-p)}$

$E(\text{silver}) = .3(40) + .5(40) = 32$

$VAR(\text{Silver}) = (40)(.3)(.7) + (40)(.5)(.5) = \sqrt{18.4}$

SD(silver) = 4.29

AP3.8 (A)



.98 wider interval →
less precise

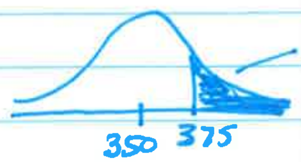
AP 3.9 (D) RELATIVE FREQUENCY – %'s because of different sample sizes

Histograms more appropriate for large data samples

AP 3.10 (C) X = 6 SIDED DIE $\mu_X = 3.5$ $\sigma_X = 1.71$ $n = 100$

$P(T \geq 375)$

?



area

$\mu_T = 3.5(100) = 350$

$\sigma_T = \frac{1.71(100)}{\sqrt{100}} = 17.1$

normalcdf (375, E99, 350, 17.1) = .0719

AP 3.11 (B) DEFINITION OF CONFOUNDING

AP 3.12 (C) CHANGING UNITS DOES NOT CHANGE CORR COEF (r)

AP 3.13 (C) Green Sheet CI: statistic ± critical value • SD statistic

$\hat{p}_1 = 88/200 = .44$

$\hat{p}_2 = 141/300 = .47$

$\hat{p}_1 - \hat{p}_2 \pm 1.96 \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$

$\Rightarrow \boxed{(.44 - .47) \pm 1.96 \sqrt{\frac{.44(.56)}{200} + \frac{.47(.53)}{300}}}$

AP 3.14 (D) BINOMIAL SETTINGS

B BINARY

I INDEPENDENT

N FIXED SAMPLE SIZE

S FIXED PROBABILITY OF SUCCESS

CLT DOES NOT APPLY TO PROPORTIONS!

AP 3.15 (A) TO HAVE A VALID EXPERIMENT THEY SHOULD HAVE PLANTED IN BOTH GARDENS

AP 3.16 (E) TYPE 2 ERROR – FAIL TO REJECT H_0 WHEN H_0 IS FALSE (H_A TRUE)

MEMORIZE:

TRUTH ABOUT THE POPULATION

Conclusion Based on sample		H_0 TRUE	H_0 FALSE (H_A TRUE)
	Reject H_0	TYPE I ERROR	POWER (correct conclusion)
	FAIL TO REJECT H_0	Correct conclusion	Type II ERROR

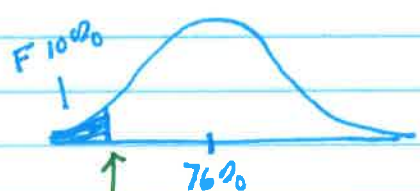
AP 3.17 (B) Skewed data use median and IQR
symmetric data use mean and std. deviation

AP 3.18 (B) DEFINITION OF A CONFIDENCE LEVEL (pg 474)

AP 3.19 (E) Corr Coff $r = .96$ measures the linear association
$r^2 = (.96)^2 = .92$ measures the strength of the model
(E) is r^2 definition in context.

AP 3.20 (C) Grades $N(76, 8)$

Grades	%	Cum %
A	10%	10%
B	20	30%
C	40	70%
D	20	90%
F	10%	100%



Grade = 65.75 ← invNorm(.1, 76, 8)

AP 3.21 (a) median – 50%tile – largest bar is 140–180
mean – reasonable to be higher since distribution is skewed right and mean will be pulled right

AP 3.22 (D) 95% CI SRS $n = 50 \longrightarrow = 200$

p = proportion with Facebook

GREEN SHEET: CI: Statistic ± $\underbrace{\text{Critical Value} * \text{S.D.}}_{\text{Margin of Error}}$

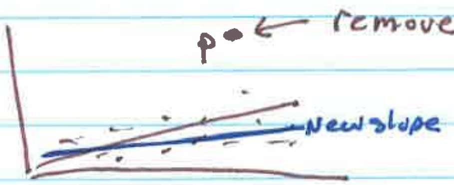
SD of sample proportion

$$\sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{p(1-p)}{50}} \Rightarrow \sqrt{\frac{p(1-p)}{4(50)}}$$

$$\Rightarrow \frac{1}{2}\sqrt{\frac{p(1-p)}{5}}$$

Margin of Error Divided by 2

AP3.23 (B)



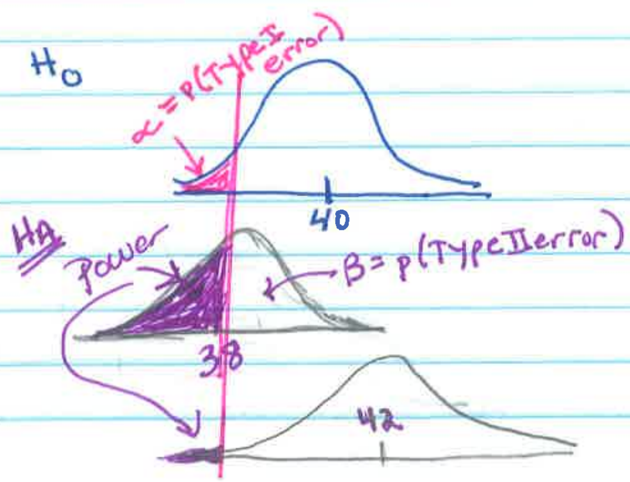
① removing the outlier will increase the association between the 2 variable which is the correlation

② With the point (P) the slope of the line is pulled towards the point. Removing the point will decrease the slope.

AP3.24 (E) Jan → variability 15 – 90 Range = 75

July → variability 55 – 95 Range = 40

AP3.25 (A)



Ho : μ = 40

Ha : μ < 40

The smaller the value, the larger the power

Pg 541

AP 3.26 (B) STRATIFIED RANDOM SAMPLE DIVIDES THE POPULATION INTO STRATA OF SIMILIAR GROUPS OF INDIVIDUALS.

KNOW THE DIFFERENCE BETWEEN Sampling Techniques – SRS

STRATIFIED VS CLUSTER

VOLUNTARY AND CONVENIENCE

$\boxed{\text{Pg } 225}$

AP 3.27 (C) $p = .45$ SRS $n = 250$

GREEN SHEET $SD(\text{proportion}) = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{.45(.55)}{250}}$

$= .03146$

AP 3.28 (D) $\mu = 9$ SRS $n = 28$ $\bar{x} = 7.9$ $S_x = 2.1$

$\boxed{H_0 : \mu = 9 \qquad H_A : \mu < 9}$

AP 3.29 (A) 1 sample t-test for μ * t-test since pop SD is unkown

GREEN SHEET : $\text{Test statistic} = \frac{\text{STATISTIC} - \text{PARAMETER}}{\text{SD STATISTIC}}$

AND SD OF a Sample mean $= \frac{\sigma}{\sqrt{n}}$

$\rightarrow t = \frac{7.9 - 9}{2.1/\sqrt{28}}$

AP 3.30 (B)

OBama		1029
McCain		695
Other		21
Undecide		105
		1850

Marginal distributions

Marginal Distributions

Obama $1029/1850 = .556$

McCain $695/1850 = .3757$

etc.

AP3.31

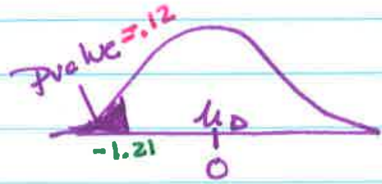
TEST: 1 sample t-test for difference of mean

μ_D = Weight Loss NOT REGAINED (AFTER−BEFORE)

CALC
L1 = Before
L2 = After
L3 = L2−L1
STAT> TESTS>
2: TTEST

$H_0: \mu_D = 0$
$H_A: \mu_D < 0$

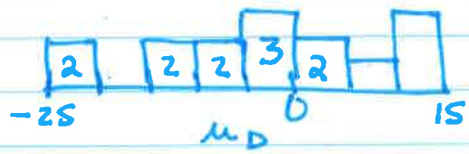
$\alpha = .05$



Paired t-test b/c same dieters receive both treatments

TIP: P(A) ≠ P(B)

CONDITIONS
Random: random sample of dieters
Independent: a well designed experiment so we assume the difference in weights for each dieter is independent
Normal: the sample is small so we must look at a graph of the data. Look at a box plot shows No outliers; and the histogram shows no skewness



$$t = \frac{\bar{x}_d - \mu_d}{s_d/\sqrt{n}} = \frac{-3.6-0}{11.52/\sqrt{15}} = -1.21$$

$\bar{x}_d = -3.6$ $n = 15$
$s_x = 11.52$ $df = 14$

Pvalue = $P(t \le -1.21)$ = .1232

Since the pvalue is large and greater than $\alpha = .05$, we fail to reject H_0. We do not have convincing evidence that the diet has a long term effect and it appear dieters were not able to keep the weight off.

AP 3.32 (a) This is an observational study because there were NO TREATMENTS imposed

(b)

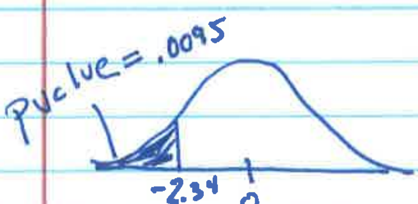
	n	# graduate HS	%
low birth weight	242	179	.7397
Control group	233	193	.8283

p_1 = true proportion HS graduates for low birth weight
p_2 = true proportion of HS graduates for control group

$H_0: p_1 = p_2$ $\alpha = .05$
$H_a: p_1 < p_2$

TEST: 2 sample Z test for the difference of $p_1 - p_2$

Conditions:
Random: 2 random samples for VLBW + Control groups
Independent; the 2 groups are independent
Normal: successes and failures for both groups are greater than 10.
VLBW – 179 and 63
Control – 193 and 40



$\hat{p}_1 = .74$
$\hat{p}_2 = .83$

pooled $\hat{p}_c = \frac{179+193}{242+233}$
$\hat{p}_c = .78$

$Z = -2.34$
pvalue $= P(Z \leq -2.34) = .0095$

Since the pvalue is small and less than $\alpha = .05$, we reject H_0. We have convincing evidence that the proportion of H.S. graduates is less for low birth weight babies than normal weight babies

AP 3.33 (a) $\widehat{\text{distance}} = -73.64 + 5.7188(\text{temperature})$

(b) The slope = 5.7188.
FOR EVERY 1 DEGREE (celsius) increase in the water discharge temperature, the predicted increase in the distance of the nearest fish from the out flow pipe is about 5.72 meters, on average.

(c) YES. THE RESIDUAL PLOT SHOWS NO APPARANT PATTERN SO A LINEAR MODEL IS APPROPRIATE.

Remember to describe the association in a scatter plot WITH - STRENGTH, SHAPE + DIRECTION.
IN THIS EXAMPLE, THE ASSOCIATION BETWEEN TEMPERATURE AND DISTANCE IS A STRONG, POSITIVE, LINEAR ASSOCIATION.

(d) TEMP = 29°C

$\widehat{\text{distance}} = -73.64 + 5.7188(29) = 92.21\text{m}$

Looking at the residual plot, 92.21 m has a residual of about -18. THIS MEANS THAT FOR THIS TEMPERATURE THE MODEL IS OVER PREDICTING.

AP 3.34

Box of Truffles

8 Chocolate	$\bar{x} = 2oz$	$S_x = .5oz$
2 Caramel	$\bar{x} = 4oz$	$S_x = 1oz$
Box (1)	$\bar{x} = 3oz$	$S_x = .2oz$

(A) Weights of truffles and the box are independent

DEFINE A NEW Random Variable:

W = the weight of a randomly selected box

$\boxed{\mu_x = E(X) = \sum x_i p_i}$

on green sheet

$\boxed{\sigma_x^2 = VAR(X) = \sum (x_i - \mu_x)^2 p_i}$

on green sheet

See combining RV's

$\mu_W = 8(2) + 2(4) + 1(3)$

$\mu_W = 27 oz$

$VAR(W) = 8(.5)^2 + 2(1)^2 + 1(.2)^2$

$VAR(W) = \boxed{4.04}$

$\sigma_W = 2.01 oz$

(B) Assume the weights of truffles and the box are approximately NORMAL → $N(27, 2.01)$

MUST STATE CLEARLY FOR THIS TYPE OF PROBLEM

$P(W > 30) = .0678$

Calc Command normalcdf (30, E99, 27, 2.01)

← Draw Graph IF YOU DO NOT FIND Z SCORE

Method 2:

$P(Z > \frac{30-27}{2.01} = 1.49)$

$= .0681$

normalcdf (1.49, E99, 0, 1)

ANSWER IN CONTEXT: There is about a 6.8% chance of randomly selecting a box of chocolates that weighs more than 30 oz.

AP 3.34 (CONT)

(C) Randomly selecting at least 1 out of 5 boxes weighing more than 30 oz's

P(at least 1 box > 30 oz) = 1 − P(none of the boxes > 30 oz)

ATLEAST PROBABILITIES FIND BY FINDING 1 − P(NONE)

P(W > 30) = .0678 (from b)

P(NONE) = 1 − .0678 = .9322

5 boxes

P(at least 1 box in 5 > 30 oz) = $1 - (.9322)^5$

1 − .7040 = .2960

ANSWER IN CONTEXT: There is about a 29.6% chance that at least 1 of 5 randomly selected box weighs more than 30 ounces.

(D) Randomly selecting 5 boxes with mean weight greater than 30 oz. This is a sampling distribution

* DEFINE VARIABLE OF INTEREST $\bar{X}$ = mean weight of 5 randomly selected boxes

Green Sheet

$\mu_{\bar{x}} = \mu$

$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

$\mu_{\bar{w}} = \mu_w = 27\text{oz}$

$\sigma_{\bar{w}} = \frac{\sigma_w}{\sqrt{n}} = \frac{2.01}{\sqrt{5}} = .899\text{oz}$

NOTICE A SMALLER SD FROM Randomly selecting 1 box in (B)

STATE THE NEW DISTRIBUTION ⟶ N(27, .899)

(NORMAL SINCE UNDERLYING DISTRIBUTION WAS NORMAL)

Calc normalcdf (30, E99, 27, .899)

$P(\bar{w} > 30) = .0004$

ZScore Method: $P(z > \frac{30-27}{.899} = 3.34) = .0004$

Calc normalcdf (3.34, E99, 0, 1)

Draw Graph if you do NOT find Z.

(ANSWER IN CONTEXT) There is a .0004 probability of randomly selecting 5 boxes that have a mean weight of more than 30 oz's

IMPORTANT: UNDERSTAND WHY PROB'S ARE SO MUCH DIFFERENT BETWEEN (B) and (D)

AP 3.35 (A) NOTE: THESE RETURN RATES ON STOCKS AND NOT PROPORTIONS!

TEST: 2 SAMPLE T-TEST FOR THE DIFFERENCE OF MEANS $(\mu_1 - \mu_2)$

$\alpha = .05$

μ_A = IS THE TRUE MEAN RETURN RATE FOR STOCK A
μ_B = IS THE TRUE MEAN RETURN RATE FOR STOCK B

$H_0: \mu_A = \mu_B$
$H_A: \mu_A \neq \mu_B$

Sampling Distributions

STOCK A	STOCK B
$n = 50$	$n = 50$
$\bar{X}_A = 11.8$	$\bar{X}_B = 7.1$
$S_A = 12.9$	$S_B = 9.6$

$df = 49$

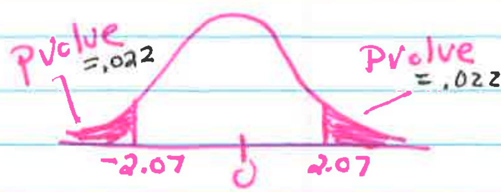
CONDITIONS
① Random samples of both stocks
② INDEPENDENT - IT IS REASONABLE THERE ARE MORE THAN 50(10) = 500 days for daily RATES IN THE PAST 5 YEARS PLUS THE 2 TYPES OF STOCKS ARE INDEPENDENT
③ Normal - the samples are sufficiently large
$n_A = n_B = 50 \geq 30$

Green Sheet
SD for diff. of means = $\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

$$t = \frac{(11.8 - 12.9) - 0}{\sqrt{\frac{(7.1)^2}{50} + \frac{(9.6)^2}{50}}} = 2.07$$

Pvalue = $P(t \leq -2.07)$ OR $P(t \geq 2.07) = .042$

Calculator Give probability from both tails



CALC tcdf(2.07, E99, 49 df)

CALC
STAT > TESTS >
2 Samp T Test
ALWAYS!
Pooled (NO)

Because the pvalue (.042) is less than $\alpha = .05$, we reject Ho. We have convincing evidence that the mean annualized returns for the 2 stocks (A and B) is different

AP 3.35 (CONT)

(B) This is basically a Question 6. They typically start with something you have been taught and then want you to use what you have been taught on a new type of problem never seen.

Do NOT Get Nervous!! You can start the Q6 and then go back to it after you complete Q's 1-5. ☺

TOH: σ_A = TRUE STANDARD DEVIATION OF STOCK A
σ_B = TRUE STANDARD DEVIATION of Stock B

$H_0: \sigma_A = \sigma_B$
$H_A: \sigma_A > \sigma_B$

(C) A NEW TEST STATISTIC $F = \frac{\text{larger sample Variance}}{\text{Smaller Sample Variance}}$

UNDERSTANDING THE NEW STATISTIC

$\sigma_A = 12.9$
$\sigma_B = 9.6$

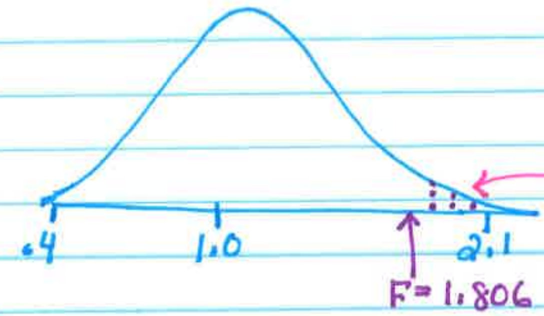
We can see that the standard deviation for stock A is greater than stock B. This would also be true for the Variances. IF THE VARIANCES WERE EQUAL THEN THE NEW TEST STATISTIC WOULD BE 1. Hence the large the value for "F" would indicate a significant difference indicating the price volatility for stock A is HIGHER THAN STOCK B.

(D) $F = \frac{(12.9)^2}{(9.6)^2} = 1.806$

AP 3.35 (CONT)

E

$\boxed{F = 1.806}$



Simulations
$n = 200$
6 were larger than our F = 1.806

$$\text{Pvalue} = P(F > 1.806) = \frac{6}{200} = \underline{.03}$$

Since the pvalue (.03) is small than $\alpha = .05$, We reject H_0. THERE IS CONVINCING EVIDENCE THAT THE TRUE STANDARD DEVIATION FOR STOCK A IS GREATER THAN THE STANDARD DEVIATION FOR STOCK B. INDICATING THE PRICE VOLATILITY FOR STOCK A IS GREATER THAN STOCK B.