Chapter 10 Review Sheet

KEY

Use the following information for questions 1-3:
Do boys perform better in math than girls? A randomly selected group of each gender were given the same math assessment. The score results in the sample were:

	Boys	Girls
n	110	135
Mean	71.6	68.3
Standard Deviation	10.4	11.2

1. What would be the null and alternative hypotheses of a test to determine if boys' scores were higher than girls' scores?
   (a) H_0: $\mu_b - \mu_g = 0$, H_a: $\mu_b - \mu_g < 0$
   (b) H_0: $\mu_b - \mu_g = 0$, H_a: $\mu_b - \mu_g \neq 0$
   (c) H_0: $\mu_b - \mu_g = 0$, H_a: $\mu_b - \mu_g > 0$
   (d) H_0: $\mu_b - \mu_g < 0$, H_a: $\mu_b - \mu_g = 0$
   (e) H_0: $\mu_b = \mu_g$, H_a: $\mu_b \neq \mu_g$

2. What procedure would you use for this test? Two-sample t test

3. Suppose the P-value of the test is .0344. We can then conclude:
   (a) At $\alpha = .025$, reject H_0.
   (b) At $\alpha = .02$, reject H_0.
   (c) At $\alpha = .01$, reject H_0.
   (d) At $\alpha = .025$, fail to reject H_0.
   (e) No conclusion can be drawn from this information.

4. There are two common methods for measuring the concentration of a pollutant in fish tissue. Do the two methods differ, on average? You apply both methods to each fish in a random sample of 18 carp and use:
   (a) the paired t test for μ_d.
   (b) the one-sample z test for p.
   (c) the two-sample t test for $\mu_1 - \mu_2$.
   (d) the two-sample z test for $p_1 - p_2$.
   (e) none of these.

oops! Meant to have this one with the situation in #1-3. Sorry for any confusion!

5. Which of the following describes a Type II error in the context of this study?
   (a) Finding convincing evidence that the true means are different for males and females, when in reality the true means are the same.
   (b) Finding convincing evidence that the true means are different for males and females, when in reality the true means are different.
   (c) Not finding convincing evidence that the true means are different for males and females, when in reality the true means are the same.
   (d) Not finding convincing evidence that the true means are different for males and females, when in reality the true means are different.
   (e) Not finding convincing evidence that the true means are different for males and females, when in reality there is convincing evidence that the true means are different.

6. A researcher reports that 80% of high school graduates, but only 40% of high school dropouts, would pass a basic literacy test. Assume that the researcher's claim is true. Suppose we give a basic literacy test to a random sample of 60 high school graduates and a separate random sample of 75 high school dropouts. Let $\hat{p}_G$ and $\hat{p}_D$ be the sample proportions of graduates and dropouts, respectively, who pass the test.

a) What is the shape of the sampling distribution of $\hat{p}_G - \hat{p}_D$?

Approximately Normal since $60(.8) = 48 \geq 10$ $\quad 60(.2) = 12 \geq 10$

(Large Counts Condition) $\quad 75(.4) = 30 \geq 10$ $\quad 75(.6) = 45 \geq 10$

b) Find the mean of the sampling distribution.

$$\mu_{\hat{p}_G - \hat{p}_D} = p_G - p_D = .8 - .4 = \boxed{.4}$$

c) Find the standard deviation of the sampling distribution.

$$\sigma_{\hat{p}_G - \hat{p}_D} = \sqrt{\frac{p_G(1-p_G)}{n_G} + \frac{p_D(1-p_D)}{n_D}} = \sqrt{\frac{.8(.2)}{60} + \frac{.4(.6)}{75}} = \boxed{.0766}$$

10% Condition: There are at least 600 h.s. graduates and 750 h.s. dropouts.

7. The heights of young men follow a Normal distribution with mean 69.3 inches and standard deviation 2.8 inches. The heights of young women follow a Normal distribution with mean 64.5 inches and standard deviation 2.5 inches. Suppose we select independent SRSs of 16 young men and 9 young women and calculate the sample mean heights $\bar{x}_M$ and $\bar{x}_W$.

a) What is the shape of the sampling distribution of $\bar{x}_M - \bar{x}_W$?

Normal since both populations are Normal.

b) Find the mean of the sampling distribution.

$$\mu_{\bar{x}_M - \bar{x}_W} = \mu_M - \mu_W = 69.3 - 64.5 = \boxed{4.8}$$

c) Find the standard deviation of the sampling distribution.

$$\sigma_{\bar{x}_M - \bar{x}_W} = \sqrt{\frac{\sigma_M^2}{n_M} + \frac{\sigma_W^2}{n_W}} = \sqrt{\frac{2.8^2}{16} + \frac{2.5^2}{9}} = \boxed{1.0883}$$

10% Condition: There are at least 160 young men and 90 young women.

8. A person released from prison before completing the original sentence is placed under the supervision of a parole board. If that person violates specified conditions of good behavior during the parole period, the board can order a return to prison. The article "Impulsive and Premeditated Homicide: An Analysis of the Subsequent Parole Risk of the Murderer" reported the data on parole behavior below. One random sample of individuals had served time in prison for impulsive murder, and the other random sample had served time for premeditated murder. Construct and interpret a 98% confidence interval for the difference in the proportions of individuals serving time for impulsive murder and individuals serving time for premeditated murder who successfully completed parole.

	Impulsive	Premeditated
Sample size	42	40
Number with no violation	13	22
Sample proportion	.310	.550

P: p_1 = proportion of all individuals serving time for impulsive murder who successfully completed parole.

p_2 = proportion of all individuals serving time for premeditated murder who successfully completed parole.

We are interested in estimating $p_1 - p_2$.

A: ✓Random - Independent random samples were taken.

✓Normal - Approximately Normal since $42(.31) = 13 \geq 10$ $\quad 42(.69) = 29 \geq 10$

$40(.55) = 22 \geq 10$ $\quad 40(.45) = 18 \geq 10$

✓Independent - $N_1 \geq 420$, $N_2 \geq 400$ It's safe to assume there are at least 420 individuals serving time for impulsive murder and at least 400 individuals serving time for premeditated murder.

N: Two-proportion z interval

I: $$(\hat{p}_1 - \hat{p}_2) \pm z^* \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} \rightarrow (.31 - .55) \pm 2.326 \sqrt{\frac{(.31)(.69)}{42} + \frac{(.55)(.45)}{40}}$$

$(-.4870, .0070)$ by hand OR $(-.4875, .00656)$ by calculator

C: We are 98% confident that the interval from −48.75% to 0.66% captures the true difference between the proportions of individuals serving time for impulsive murder and individuals serving time for premeditated murder who successfully completed parole.

9. A study of iron deficiency in infants compared independent random samples of infants whose mothers chose different ways of feeding them. One group contained breast-fed infants. The children in another group were fed a standard baby formula without any iron supplements. Here are summary results on blood hemoglobin levels at 12 months of age. Graphical displays of the sample data show no strong skewness and no outliers.

Group	n	$\bar{x}$	s
Breast-Fed	23	13.3	1.7
Formula	19	12.4	1.8

Is there significant evidence at the $\alpha = 0.05$ level that the mean hemoglobin level is different among breast-fed babies?

P: μ_B = mean hemoglobin level of all breast-fed babies.
μ_F = mean hemoglobin level of all formula-fed babies.

H: $H_0: \mu_B - \mu_F = 0$ mean hemoglobin levels are the same among breast-fed + formula babies.
$H_a: \mu_B - \mu_F \neq 0$ mean hemoglobin levels are different among breast-fed + formula babies.

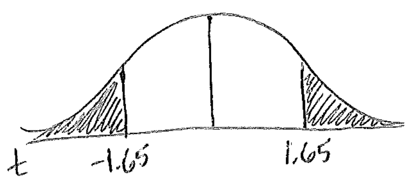
A: ✓Random – Independent random samples were taken.
✓Normal – Graphical displays of the sample data show no strong skewness + no outliers.
✓Independent – $N_B \geq 230$ There are at least 230 breast-fed babies and at least
$N_F \geq 190$ 190 formula-fed babies.

N: Since σ_B and σ_F are unknown, we will use a two-sample t test.

T: $t = \dfrac{(\bar{x}_B - \bar{x}_F) - 0}{\sqrt{\dfrac{s_B^2}{n_B} + \dfrac{s_F^2}{n_F}}} = \dfrac{13.3 - 12.4}{\sqrt{\dfrac{1.7^2}{23} + \dfrac{1.8^2}{19}}} = \dfrac{.9}{.5442} = 1.65$

df = 18 (conservative)
OR
df = 37.6 (calculator)

O: P-value = $P(t > 1.65) \times 2$ = between .10 and .20 (table)
= .1065 (calc)



M: Because the P-value is not significant at the $\alpha = .05$ level, we fail to reject H_0.

S: There is not strong evidence that the mean hemoglobin level in breast-fed babies is different than in formula-fed babies.

10. Patients with heart-attack symptoms arrive at an emergency room either by ambulance or self-transportation provided by themselves, family of friends. When a patient arrives at the emergency room, the time of arrival is recorded. The time when the patient's diagnostic treatment begins is also recorded. An administrator of a large hospital wanted to determine whether the mean wait time (time between arrival and diagnostic treatment) for patients with heart-attack symptoms differs according to the mode of transportation. Independent random samples of 75 patients with heart-attack symptoms who had reported to the emergency room via each mode of transportation were selected. For each patient, the wait time was recorded. Summary statistics for each mode of transportation are shown in the table below.

Mode of Transportation	Sample Size	Mean Wait Time (in minutes)	Standard Deviation of Wait Times (in minutes)
Ambulance	75	6.04	4.30
Self	75	8.30	5.16

a) Construct and interpret a 99% confidence interval to estimate the difference between the mean wait times for ambulance-transported patients and self-transported patients at this emergency room.

P: μ_A = mean wait times for <u>all</u> ambulance-transported patients at this ER.
μ_S = mean wait times for <u>all</u> self-transported patients at this ER.
We're interested in estimating $\mu_A - \mu_S$.

A: ✓Random - Independent random samples were taken.
✓Normal - Approximately Normal by the CLT since $n_A = 75 \geq 30$ and $n_S = 75 \geq 30$.
✓Independent - $N_A \geq 750$, $N_S \geq 750$ It's safe to assume an ER sees at least 750 patients who are transported by ambulance and at least 750 by self.

N: Since σ_A and σ_S are unknown, we will use a two-sample t interval.

I: $(\bar{x}_A - \bar{x}_S) \pm t^* \sqrt{\frac{s_A^2}{n_A} + \frac{s_S^2}{n_S}} \rightarrow (6.04 - 8.30) \pm 2.660 \sqrt{\frac{4.30^2}{75} + \frac{5.16^2}{75}} \rightarrow (-4.32, -.197)$

df = 74 (conservative)
↳ df = 60 on table

OR
↳ $(-4.285, -.2353)$ df = 143.34 (calculator)

C: We are 99% confident that the interval from −4.285 to −.2353 captures the true difference in wait times for ambulance and self-transported patients at this ER.

b) Based only on this confidence interval, is there a significant difference in mean wait times? Justify your answer.

Yes, because 0 is not in the interval. It is significant at the 1% level.