Medical Stats Practice Final exam ~ Explanations (Q#1-11)
*(*All referenced "Final Reviews" refer to the TA Mark Nessim's final review notes (Parts 1-3))*

1. Reference: Lec. 18-19, *also make sure to study different examples of scatter plots in this lecture with different r (corr. coeff.) values…Dr. Douglas said they're easy to know for the final!)*
   -Answer: c
   -When correlation coefficient, paired data set. (eg: # states data paired with their cancer mortality rates). Therefore, n will account for both data sets. → n = 7
   -Find means for each data set: $\bar{x} = 8.14$, $\bar{y} = 8.86$
   -Know that correlation coefficient: $-1 < r < 1$
   -Calculate using formula:

$$\text{Correlation coefficient: } r = \frac{cov(x,y)}{S_x S_y} = \frac{cov(\text{Data set 1, Data set 2})}{S_1 S_2}$$

$$\boxed{r = \frac{\Sigma x_i y_i - n\bar{x}\bar{y}}{\Sigma x_i^2 - n\bar{x}^2}}$$

Need: $\Sigma x_i y_i = (5)(9) + (7)(14) \cdot + (14)(4) = 606$

$\Sigma x_i^2 = (5)^2 + (7)^2 + \cdots (14 = 641$

$$= \frac{(606) - (7)(8.14)(8.86)}{(641) - (7)(8.14)^2}$$

$$= \frac{101.1572}{177.1828}$$

$$\boxed{r = 0.571}$$

(closest to answer c)

$\bar{x}$ (Data set #1) = 8.14
$\bar{y}$ (Data set #2) = 8.86
n = 7

2. Reference: Lec 18-19; Final Review, Part 2, Slide #6
   -Answer: b
   - "The best line is the one that minimizes the sum of squared vertical differences between the points and the line." (remember, "least *squares* fit")

---

3. Reference: Lec 18-19
   -Answer: b
   -Sample correlation coefficient (*simple linear regression*):
   r = 1 strong (+) linear relationship
   r = 0 no linear relationship
   r = -1 strong (-) linear relationship
   -Sample correlation coefficient for this problem (given): r = 0.960702626…close to 1.0 so strong (+) linear relationship

4. Reference: Lec 18-19
   -Answer: d
   -problem asking to test for *positive linear relationship*, so based on above, know that want slope (β_1) to be positive. Therefore want H_1: $\beta_1 > 0$. Only one option → answer d

5. Reference: Lec 18-19; Final Review, Part 2, Slide 10-12, 15-16
   -Answer: c

$$\text{Test statistic} = \frac{\hat{\beta} - \beta_0}{S_b} \quad \text{OR} \quad \frac{b_1 - \beta_1}{S_{b_1}} \quad \text{(same thing)}$$

b_1 = coeff. of x-var (slope) = 0.0949 [given in prob.]

$\beta_1 = 0$ (see Q#4's answer)

S_{b_1} = standard error of x-var. = 0.00827 [given]

$$\boxed{\text{Test stat} = \frac{b_1 - \beta_1}{S_{b_1}}} = \frac{0.0949 - 0}{0.00827} = \boxed{11.475}$$ (closest to answer c

6. Reference: Lec 18-19; Final Review, Part 2, Slide 10-12, 15-16
   -Answer: b
   -Looking for p-value that is "strictly <u>greater than zero</u>". Normally calculate p-value for both tails of a bell-shaped curve. However, in this problem, only want it for values greater than 0, so only account for the right half of the curve. Therefore, take ½ of the given p-value of the X-var. (degrees) → answer b.

7. Reference: Lec 18-19
   -Answer: a
   -Based on data given → regression line: y = 9.445 + 0.0949x
   -slope = 0.0949 → (+) value so know that for each 1 unit (in this case, of temperature), y-value should *increase* ("more")
   -Note!: data said that it's "in thousands of people" SO:
   (heat exhaustion rate)(1000 people) = (0.0949)(1000) = 94.9 ≅ 95 people.
   -Given all this, answer a fits that for ever 1 degree rise in temperature, 95 *more* people suffer from heat exhaustion.

---

8. Reference: Lec 13-14
   -Answer: e (none of the above)
   -Wrong answers:
   a: No, these hypotheses refer to double-sided test. Problem calls for one-sided test "more expensive than". If it was a double-sided test, it would call for a set value, and the alternate hypothesis would be if the situation was that it was NOT that set value.
   b: False. Calculate t-test with polled variance (s^2_p) from the 2 sample variances
   c: No, for case of "normal populations with equal variances" (pay attention to initial problem description) with small sample sizes, use t-distribution with df = $n_1 + n_2$ -2.
   This case: df = 20 + 20 – 2 = 38
   d: No. this is what you would want to do (calculate the pooled variance)

9. Reference: Lec 13-14
   -Answer: a
   -Problem asking for "sample difference in means" → $\overline{X_1} - \overline{X_2}$
   -Population 1 (Viagra) $\overline{X_1} = 84.11$
   Population 2 (Levitra) $\overline{X_2} = 99.20$
   → $\overline{X_1} - \overline{X_2} = 84.11 - 99.20 = -15.09$

10. Reference: Lec 13-14
    -Answer: d

$H_0: \bar{x}_2 \text{ (Levitra)} - \bar{x}_1 \text{ (Viagra)} = 10$

$H_1: \bar{x}_2 - \bar{x}_1 \neq 10$

$\alpha = 0.10 \rightarrow$ 2-sided test, check for 0.05 on t-table

if $\bar{x}_2 - \bar{x}_1 = 10 \rightarrow \bar{x}_1 - \bar{x}_2 = -10$

$$\text{Test stat} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_p^2/n_1 + s_p^2/n_2}} = \frac{-10}{\sqrt{\frac{79.80}{20} + \frac{79.80}{20}}} = \boxed{-3.54}$$

$t_{0.05, 38} = 2.024$

$|-3.54| > 2.024 \therefore$ Reject H_0!

Viagra (1)	Levitra (2)
$n_1 = 20$	$n_2 = 20$
$\bar{x}_1 = 84.11$	$\bar{x}_2 = 99.20$
$s_1 = 7.04$	$s_2 = 10.49$

Pooled variance (s_p^2)

$$\boxed{s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}$$

$$= \frac{(19)(7.04)^2 + (19)(10.49)^2}{20+20-2}$$

$s_p^2 = 79.80$

11. Reference: Final Review, Part 1, Slide #27
    -Answer: e (all of the above)
    -see referenced slide. Spells it out for each of the different answers (which are all correct characteristics of *Ratio of 2 Variances* test)