Use the following information to answer questions 1-4.

Consider the population of female diabetics over 30. A sample of 25 diabetic women is selected, with mean $\bar{x} = 84$ mmHg and standard deviation $s = 12.1$ mmHg. In the general population of women over 30, it is thought that the average DBP is 74.4 mmHg.

1. Construct a 95% confidence interval for the true population mean μ of DBP for female diabetics over 30. (4 pts)

   - $\bar{x} = 84$, $\mu_0 = 74.4$
   - $df = 25 - 1 = 24 \;\Rightarrow\; t_{0.975,24} = 2.064$
   - $SE_{\bar{x}} = \dfrac{s}{\sqrt{n}} = \dfrac{12.1}{\sqrt{25}} = \dfrac{12.1}{5} = 2.42$

   $\Rightarrow$ 95% CI: $\bar{x} \pm t_{df,1-\alpha/2} \cdot SE_{\bar{x}} = 84 \pm 2.064(2.42) = 84 \pm 4.995 = (79.005, 88.995)$

2. List the null and alternative hypotheses for the hypothesis test corresponding to the CI you created in #1. (2 pts)

$$H_0 : \mu = 74.4$$
$$H_A : \mu \neq 74.4$$

3. *Without conducting a formal hypothesis test*, what is your decision about this test regarding the null hypothesis? (No interpretation necessary.) (2 pts)

   We see that $\mu_0 = 74.4$ is NOT in our 95% CI, therefore we would reject H_0.

4. How large a sample would be required for the 95% confidence interval to have a margin of error of 3.5 mmHg? (4 pts)

$$n = \left(z_{0.975} \cdot \frac{\sigma}{m}\right)^2 = \left(1.96 \cdot \frac{12.1}{3.5}\right)^2 = 6.776^2 = 45.91$$

   Rounding up, we take $n = 46$.

Use the following information to answer questions 5-8.

E. canis infection is a tick-borne disease of dogs that is sometimes contracted by humans. Among infected humans, the distribution of white blood cell counts is approximately Normal and has an unknown mean μ and a standard deviation σ. In the general population, the mean white blood cell count is 7250 mm^3.

A simple random sample of 15 infected persons is used to test if the average WBC count differs from that of the general population. The sample mean and standard deviation are $\bar{x} = 4767$ mm^3 and $s = 3204$ mm^3. A 2-sided one sample t-test will be conducted at $\alpha = 0.05$.

__C__ 5. What are the appropriate degrees of freedom for this test? (3 pts)

A) 15
B) 7249
C) 14
D) 4767

__B__ 6. What is the p-value for this t test? (3 pts)

A) $0.0025 < p\text{-value} < 0.005$
B) $0.005 < p\text{-value} < 0.01$
C) $0.01 < p\text{-value} < 0.025$
D) $p\text{-value} > 0.5$

__B__ 7. Suppose the mean and standard deviation obtained were based on a sample of size $n = 30$ infected persons rather than 15. What do we know about the value of the p-value? (3 pts)

A) It would be larger.
B) It would be smaller.
C) It would be unchanged, because the detectable difference Δ is unchanged.
D) It would be unchanged, because the variability measured by the standard deviation s stays the same.

__D__ 8. If we decrease α, what happens to the p-value and the power of the test? (3 pts)

A) p-value will increase and power will increase.
B) p-value will not change and power will not change.
C) p-value will decrease and power will decrease.
D) p-value will not change and power will decrease.
E) p-value will decrease and power will increase.

Use the following information to answer questions 9-12.

Consider a study comparing the characteristics of tuberculosis meningitis in patients infected with HIV and those who are not infected. We would like to determine whether the two populations have the same mean age.

Group	Mean ($\bar{x}_i$)	SD (s_i)	n_i
Group 1 (HIV Negative)	38.8	21.7	19
Group 2 (HIV Positive)	27.9	5.6	37

9. Assuming unequal group variances, what is the standard error of the difference between the means? (4 pts)

$$SE_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}} = \sqrt{\frac{21.7^2}{19}+\frac{5.6^2}{37}} = \sqrt{24.7837+0.8476} = \sqrt{25.6313} = 5.0627$$

10. What is the test statistic for the two-sided test of no mean difference ($H_0 : \mu_1 = \mu_2$)? (4 pts)

$$t = \frac{\bar{x}_1-\bar{x}_2}{SE_{\bar{x}_1-\bar{x}_2}} = \frac{38.8-27.9}{5.0627} = 2.153$$

11. Using the conservative degrees of freedom, what is the corresponding p-value? (4 pts)

Using Table C with $df = 18$, we see that

$$2.101 < t < 2.552$$

$$0.01 < p_{\text{1-sided}} < 0.025$$

$$0.02 < p_{\text{2-sided}} < 0.05$$

___F___ 12. **True/False.** Assume the 95% confidence interval for $\mu_1-\mu_2$ is given by (0.263,21.537). Then the true (population) mean difference $\mu_1-\mu_2$ has a 95% probability of being between 0.263 and 21.537. (2 pts)

This problem concerns the definition of the confidence interval.

The population mean difference $\mu_1-\mu_2$ is a constant and therefore does not have any probability associated with it! The probability is associated with the confidence interval, and its likelihood of capturing the true mean difference $\mu_1-\mu_2$.

Use the following information to answer questions 13-14.

A paired-samples study looked at the number of cavity-free children (per 100) before and after public water fluoridation projects in $n = 16$ U.S. cities. After calculating the differences, we obtain $\bar{x}_d = 12.2$ and $s_d = 13.62$. Based on the data, we decide to reject $H_0 : \mu_d = 0$ in favor of $H_A : \mu_d \neq 0$ at the $\alpha = 0.05$ level of significance.

___B___ 13. Suppose the difference worth detecting $\Delta = 10$. What is the power of the test, assuming $\sigma = s_d = 13.62$? (5 pts)

A) 0.6368

B) 0.8365

C) 0.9726

D) 0.1635

___A___ 14. Suppose we want to decrease the difference worth detecting to $\Delta = 5$. What happens to the Type II error rate of the test? (2 pts)

A) Type II error rate increases.

B) Type II error rate decreases.

C) Type II error rate is unchanged.

___F___ 15. **True/False.** When the original population is Normal, you can only use t procedures on samples of large size ($n > 40$). (3 pts)

___F___ 16. **True/False.** ANOVA tests whether the variances of k groups are the same. (3 pts)

___T___ 17. **True/False.** ANOVA requires that individual observations are independent of each other. (3 pts)

___T___ 18. **True/False.** ANOVA requires that group samples are independent of each other. (3 pts)

19. Complete the following ANOVA table. (10 pts)

Source	DF	Sum of Squares	Mean Square	F-Value	Pr > F
Model (Between)	3	3.456	1.152	7.68	0.0097
Error (Within)	8	1.2	0.15		
Corrected Total	11	4.656			

Use the following information to answer questions 21-23.

A trial evaluated the fever-reducing effects of three treatment therapies (aspirin, ibuprofen and acetaminophen). Study subjects were 15 adults seen in an emergency room with diagnoses of flu with body temperatures between 100.0°F to 100.9°F and were randomly assigned to one of the treatment groups. The response variable is the resulting decrease in body temperature, two hours post-therapy. The ANOVA table is provided below.

Source	DF	Sum of Squares	Mean Square	F-Value	Pr > F
Model (Between)	2	3.432	1.716	4.789	0.0296
Error (Within)	12	4.300	0.358		
Corrected Total	14	7.731			

20. Write the null and alternative hypotheses that are tested by the F statistic. (2 pts)

$$H_0 : \mu_1 = \mu_2 = \mu_3$$
$$H_A : \text{at least one of the } \mu_i\text{'s differs.}$$

___A___ 21. Which distribution was used to find the P-value? (3 pts)

A) F with 2 and 12 degrees of freedom.
B) F with 2 and 14 degrees of freedom.
C) F with 12 and 14 degrees of freedom.
D) t with 12 degrees of freedom.

___D___ 22. What is the value of the estimated pooled standard deviation σ? (3 pts)

A) 2.074
B) 0.358
C) 1.716
D) 0.598

___C___ 23. At a significance level of 0.05, what is the appropriate conclusion about mean fever reduction? (3 pts)

A) The average reduction in fever appears to be the same for all three therapies.
B) The average reduction in fever appears to be different for each of the three therapies.
C) It appears that at least one of the three therapies has a different average reduction in fever.
D) One therapy is significantly better than the other two.

Use the following to answer questions 24-29.

A study was conducted to monitor the emissions (in μm^3) of a noxious substance from a chemical plant and the concentration of the chemical at a location in close proximity to the plant at various times throughout the year. A total of 14 measurements were made. SAS output for the least squares model fit is provided (some entries have been omitted):

Fitted Model: Concentration = 1.543 + 1.825*Emissions

Summary of Fit Table

R-Square	R-Square Adj.	Root MSE	Mean of Response
0.7939	0.7767	1.5140	8.8107

ANOVA Table

Source	DF	Sum of Squares	Mean Square	F	Pr> F
Between (Model)		105.9639		46.2294	< .0001
Within (Error)	12				
Total		133.4695			

Parameter Estimate Table

Variable	Estimate	Std Error	t Value	Prob> $\|t\|$
Intercept	1.543	1.1429		0.2019
Emissions	1.825	0.2684		< .0001

C 24. The degrees of freedom for SS_B and SS_W are, respectively (3 pts)

A) $df_B = 2$, $df_W = 12$.

B) $df_B = 2$, $df_W = 13$.

C) $df_B = 1$, $df_W = 12$.

D) $df_B = 1$, $df_W = 13$.

E) $df_B = 1$, $df_W = 14$.

B 25. What is the test statistic and its value to test $H_0 : \beta = 0$ against $H_A : \beta \neq 0$? (3 pts)

A) F = 2.292

B) t = 6.80

C) t = 1.35

D) F = 3.8524

___A___ 26. What is the 95% confidence interval estimate for β? (4 pts)

A) (1.24, 2.41)

B) (-0.49, 3.58)

C) (1.35, 2.38)

D) (-0.95, 4.03)

E) (-1.76, 4.84)

27. Give an interpretation of the estimate of the slope in the context of the problem, in language the investigators can understand. (4 pts)

We estimate the slope β to be $b = 1.825$. This means that for every 1-unit (μm^3) increase we see in Emissions, we expect the concentration of the chemical to increase by 1.825 units.

28. Based on this model, what is the estimated (predicted) Concentration level of the chemical, given an Emissions level of 10 μm^3? (4 pts)

$$\hat{y} = 1.543 + 1.825(10) = 19.793$$

___D___ 29. From the SPSS output, the value of R^2 equals 0.7939. Which of the following interpretations of the meaning of this value is correct? (3 pts)

A) 79.39% of the variation in the predicted values of Emissions can be explained by its linear relationship with Concentration.

B) 79.39% of the variation in Emissions can be explained by a nonlinear relationship with Concentration.

C) 79.39% of the variation in the predicted values of Concentration can be explained by its linear relationship with Emissions.

D) 79.39% of the variation in Concentration can be explained by its linear relationship with Emissions.