For questions 1 – 5: Identify the data type. Use the following options.

`A.` Nominal
`B.` Ordinal
`C.` Interval
`D.` Ratio

1. *(1 pt)* Presence of head or neck cancer (yes or no).
2. *(1 pt)* White blood cells per deciliter of whole blood.
3. *(1 pt)* Grade in BIOS 600 (H, P, L, or F).
4. *(1 pt)* Birthweight of full term male infant (grams).
5. *(1 pt)* Treatment group (Placebo, Treatment A, Treatment B).

For questions 6 – 10: Identify the data type. Use the following options.

`A.` Continuous
`B.` Discrete

6. *(1 pt)* Movie ratings (★, ★★, . . ., ★★★★★).
7. *(1 pt)* Height (cm).
8. *(1 pt)* Lung Volume (cm^3)
9. *(1 pt)* Gravidity. (Number of times a women has been pregnant.)
10. *(1 pt)* Parity. (Number of times a women has given birth.)

11. *(2 pts)* Calculate Q3 from the following data.

i	1	2	3	4	5	6	7	8	9	10
$X_{(i)}$	0.03	0.05	0.05	0.05	0.12	0.19	0.19	0.20	0.26	0.29

For questions 12 – 15: For each situation, select the graphical display that best achieves the researcher's goals. Use the following options.

A. Scatterplot
B. Histogram
C. Boxplots
D. Time Series plot
E. Bar Chart
F. Pie Chart
G. None of the above

12. *(2 pts)* Display the bimodal distribution of X

13. *(2 pts)* Display the nonlinear relationship between X and Y

14. *(2 pts)* Show the change in birth rates over time

15. *(2 pts)* Compare the median and IQR of of X for two populations

16. *(1 pt)* A fair coin is tossed 8 times. Two possible results are:

```
(i)  H T T H H H H T
(ii) H H H H H H H H
```

Which statement is true?

A. Sequence (i) is more likely
B. Sequence (ii) is more likely
C. Sequences (i) and (ii) are equally likely
D. Cannot say which sequence is more likely

For questions 17 – 24: The following data comes from the UNC Office of Institutional Research and Assessment. It represents the UNC undergraduate population for Fall 2011 semester.

	Asian	Black	White	Multiracial	Unknown	Other	Total
Male	737	584	5210	184	193	793	7701
Female	947	1152	6972	269	298	1091	10729
Total	1684	1736	12182	453	491	1884	18430

Suppose we randomly select an individual from this population. (Round your answers to 4 decimal places.)

17. *(1 pt)* What is the probability the individual is female?

18. *(1 pt)* What is the probability the individual is multiracial?

19. *(1 pt)* What is the probability the individual is female given that he or she is multiracial?

20. *(1 pt)* What is the probability the individual is multiracial given that she is female?

21. *(1 pt)* What is the probability the individual is female and multiracial?

22. *(1 pt)* What is the probability the individual is female or multiracial?

23. *(1 pt)* What is the probability the individual is black or white?

24. *(2 pts)* True or False? Race and sex are independent in this population.

For questions 25 – 30: Identify the distribution of the random variable in the following examples. Use the following options:

A. Normal
B. Binomial
C. Poisson
D. None of the Above

25. *(1 pt)* Let A be the number of new head or neck cancer cases in North Carolina in 2011.

26. *(1 pt)* Of the individuals diagnosed with head or neck cancer in North Carolina in 2011, let B be the number that survive five years.

27. *(1 pt)* Of the individuals in BIOS 600, let C be the number that earn an H or A.

28. *(1 pt)* Let D be the weight of a full term male infant. The average weight is 3800 grams, and the standard deviation is 448.

29. *(1 pt)* In a sample of 100 full term male infants, let E be the number that weight more than 4000 grams. The average weight is 3800 grams, and the standard deviation is 448.

30. *(1 pt)* Let G be an individual's grade in BIOS 600, either H, P, L, or F.

31. *(2 pts)* Referring to question 28, calculate $P(D > 4000)$?

32. *(2 pts)* Referring to question 29 and 31, calculate $P(E \leq 25)$.

For questions 33 – 37: Select the term that best describes the example or definition. Use the following options.

(A) Single Blind
(B) Double Blind
(C) Triple Blind
(D) Placebo Effect
(E) Experimenter's Bias
(F) Sampling Frame
(G) Probability Sample
(H) Nonresponse Bias
(I) Unit Nonresponse
(J) Item Nonresponse
(K) Convenience Sample
(L) Sampling Bias
(M) Systematic Sample
(N) Stratified Random Sampling
(O) Clustered Random Sample
(P) Simple Random Sample (SRS)
(Q) Completely Randomized Design
(R) Randomized Block Design
(S) Potential Confounding Variable

33. *(2 pts)* A list of units from which the sample is selected.

34. *(2 pts)* A random sample in which every k^{th} individual is selected after a randomly selected starting point.

35. *(2 pts)* A survey sample may suffer from this if survey respondents are systematically different than nonrespondents.

36. *(2 pts)* An experimental design in which study subjects are randomized separately at each recruitment hospital.

37. *(2 pts)* The type of experimental design in question 36 will ensure that study site is not this type of variable.

38. *(3 pts)* Calculate the correlation coefficient of X and Y.

X	Y
-1	-1
-2	-1
0	-2
2	2

39. *(1 pt)* Calculate the correlation coefficient of X and W.

X	Y+1
-1	0
-2	0
0	-1
2	3

40. *(1 pt)* True or False? A correlation of X and Y close to 0 indicates that X and Y are associated.

41. *(1 pt)* True or False? A correlation of X and Y close to -1 indicates strong linear association between X and Y.

42. *(1 pt)* True or False? A correlation of X and Y close to 1 indicates strong linear association between X and Y.

43. *(1 pt)* True or False? When predicting X from Y, a correlation of $r = 0.8$ is preferred to a correlation of $r = -0.8$.

44. *(1 pt)* True or False? A lurking variable, Z, may exaggerate the correlation between X and Y.

45. *(1 pt)* True or False? A lurking variable, Z, may minimize the correlation between X and Y.

■ In this class we have introduced the concept of population parameters. For example, we have discussed the population mean, μ, for continuous outcomes. And, we've discussed the population proportion, p, for binary outcomes.

There is a family of hypothesis tests dealing with μ, and there is a family of hypothesis tests dealing with p. Within each family, we can consider (a) hypotheses that deal with a single population, (b) hypotheses that directly compare 2 populations, and (c) hypotheses that compare 3 or more populations.

The table below organizes these 4 types of hypothesis tests. Each cell is labeled.

	μ	p
One Population	A	B
Two Populations	C	D
Three or More Populations	E	F

For problems 46 – 48, read the situation and identify which type of hypothesis test is appropriate.

46. *(1 pt)* Investigator C. Sandiego wishes to investigate regional height differences. She collects the height of 15 individuals from 3 regions: Coastal Carolina, Piedmont Carolina, and Appalachian Carolina. She collects data from a total of $15 \times 3 = 45$ individuals.

47. *(1 pt)* Researchers design an experiment in which workers are randomly assigned to either wear paper masks (control condition) or complete breathing equipment (experimental condition) during asbestos removal. Researchers record the number of workers that develop asbestosis during the following 5 years.

48. *(1 pt)* A researcher believes individuals with a specific genetic marker have a lower bone density than their counterparts in the general population. In general male population, total femur bone mineral density (g/cm^2) is known to be centered at 1. To test her hypothesis, the researcher collected femur bone mineral density samples from 51 individuals with the genetic marker.

49. *(1 pt)* Which test statistic(s) can apply to cell C? (Mark all that apply.)

A. t-test
B. 2 sample t-test
C. Large sample binomial proportion test
D. 2 Sample binomial proportion test
E. Pearson's χ^2 test
F. ANOVA F test

50. *(1 pt)* Which test statistic(s) can apply to cell D? (Mark all that apply.)

A. t-test
B. 2 sample t-test
C. Large sample binomial proportion test
D. 2 Sample binomial proportion test
E. Pearson's χ^2 test
F. ANOVA F test

51. *(1 pt)* Which test statistic(s) can apply to cell E? (Mark all that apply.)

    A. t-test
    B. 2 sample t-test
    C. Large sample binomial proportion test
    D. 2 Sample binomial proportion test
    E. Pearson's χ^2 test
    F. ANOVA F test

52. *(2 pts)* Return to the situation in question 47. Calculate the value of the 2 sample binomial proportion test statistic using the following data.

	Asbestosis	No Asbestosis
Experimental	64	136
Control	101	99

53. *(2 pts)* Return to the situation in question 48. Suppose the sample mean is 0.9 and the sample variance is 0.25. Calculate the value of the appropriate test statistic.

54. *(3 pts)* You are working with a researcher who has collected performance scores from 8 students from class A and 7 students from class B. The researcher would like to compare class A and class B in terms of performance. Which test statistic is appropriate?

A. Kruskal-Wallis Test
B. Binomial Proportion Test
C. Sign Test
D. t Test
E. Paired t Test
F. Wilcoxon-Signed-Rank Test
G. 2 Sample t Test
H. Wilcoxon-Rank-Sum Test

55. *(3 pts)* Investigator C. Sandiego wishes to find whether banks of certain countries have more money in their vaults. She inspects the vaults of several banks from several countries. Below are the results (in $ millions).

USA	11, 13, 9
Switzerland	15, 20, 22
Ecuador	4, 1, 2
Mauritius	11, 14, 16.33
Togo	5, 13, 11

Which test statistic is appropriate?

A. Kruskal-Wallis Test
B. Binomial Proportion Test
C. Sign Test
D. t Test
E. Paired t Test
F. Wilcoxon-Signed-Rank Test
G. 2 Sample t Test
H. Wilcoxon-Rank-Sum Test

56. *(3 pts)* The nonparametric equivalent of the one sample t-test is _____ .

A. Kruskal-Wallis Test
B. Binomial Proportion Test
C. Sign Test
D. ANOVA F Test
E. Paired t Test
F. Wilcoxon-Signed-Rank Test
G. 2 Sample t Test
H. Wilcoxon-Rank-Sum Test

57. *(1 pt)* Enjoy a point.

58. *(3 pts)* A nutrition researcher collected data from 1200 high school seniors about their dietary habits. The researcher calculated the mean level of fat intake (in calories) and the corresponding 95% confidence interval. The results where $\hat{\mu} = 747$ and (691, 803).

    How should this researcher interpret these results?

    A. There is a 95% probability that μ is between 691 and 803.
    B. 95% of the possible values for μ lie between 691 and 803.
    C. If we where to resample the population and recalculate the confidence interval several times, then we would expect 95% of confidence intervals to contain μ.
    D. None of the above.

59. *(3 pts)* Calculate a 90% confidence interval for μ with the following 5 data points: 6.0, 4.1, 8.3, 4.4, 5.1. Report the margin of error. (Margin of error is the confidence interval's half width.)

60. *(3 pts)* An investigator wants to demonstrate that a new diet would reduce diastolic blood pressure by 4 mmHg in three months. The investigator wants to design a two arm study (treatment and control) for hypertensive adult males. He is willing to risk a Type I error of 10 percent and a Type II error of 2.5 percent for a two sided test. Previous data show that the mean and standard deviation of diastolic blood pressure among hypertensive males are 95.4 mmHg and 5.6 mmHg. Calculate the study's required TOTAL sample size.

61. *(1 pt)* Continuing with the previous question, if the investigator prefers a Type I error of 5% instead, what would happen to the required sample size n?

    A. Goes up
    B. Goes down
    C. Stays the same
    D. Cannot say

For questions 62 – 67: Suppose we sample 400 individuals from UNC's Fall 2011 undergraduate population, and we record each individual's sex and race. The data is reported below.

	Asian	Black	Multiracial	Other	White
Females	27	26	7	36	142
Males	19	15	6	19	103

Our goal is to investigate association that may exist in this population between sex and race.

62. *(1 pt)* What is the reference distribution of the appropriate test statistic.
    - A. χ^2
    - B. Z
    - C. t
    - D. F
    - E. Poisson
    - F. Binomial
    - G. None of the above.

63. *(1 pt)* The reference distribution of the appropriate test statistic has _____ degree(s) of freedom.

64. *(2 pts)* Calculate the value of the test statistic.

65. *(2 pts)* Find the p-value.
    - A. $0.1 < p$
    - B. $0.05 < p \leq 0.1$
    - C. $0.01 < p \leq 0.05$
    - D. $p \leq 0.01$

66. *(2 pts)* What is test result at an $\alpha = 0.01$ significance level?
    - A. Fail to reject the null hypothesis
    - B. Reject the null hypothesis

67. *(2 pts)* Which conclusion is best?
    - A. The data suggests that the distribution of race is different between males and females in this population.
    - B. There is insufficient evidence to suggest that the distribution of race differs between males and females in this population.

For questions 68 – 74: Investigator C. Sandiego wishes to investigate regional height differences. She collects the height of 15 individuals from 4 regions: A, B, C, and D. She collects data from a total of $15 \times 4 = 60$ individuals. The incomplete ANOVA table follows:

Source	DF	SS	MS	F value
Country	A	C	121.442	E
Residuals	B	2669.09	D	

68. *(1 pt)* What is the value in the A cell?

69. *(1 pt)* What is the value in the B cell?

70. *(1 pt)* What is the value in the C cell?

71. *(1 pt)* What is the value in the D cell?

72. *(1 pt)* What is the value in the E cell?

73. *(2 pts)* The p-value of the ANOVA F test is 0.06. At the $\alpha = 0.05$ test level, the data would support which conclusion?

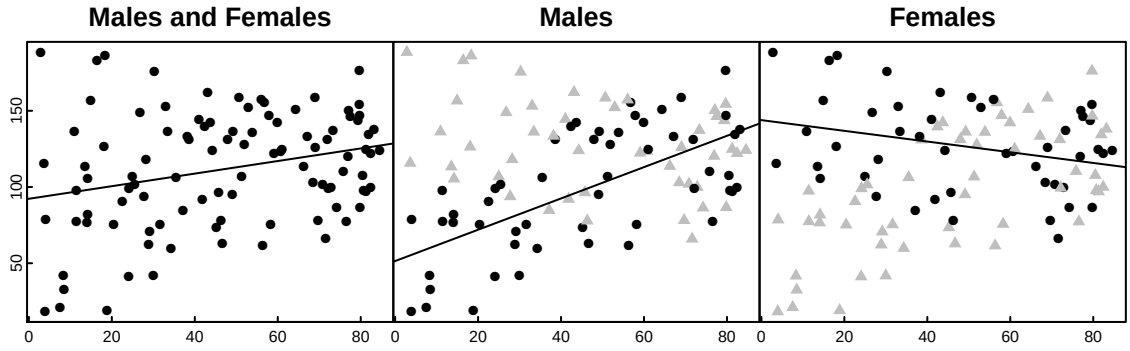
    A. The mean height in each region is the same.
    B. The mean height is different in each region.
    C. There is at least one region that differs from the rest in height.
    D. None of the above.

74. *(3 pts)* Suppose after performing the ANOVA F test, the researcher performs every pairwise comparison of the 4 regions. If the researcher performs a Bonferroni multiple comparison correction, which pairwise differences would be statistically significant? (Mark all that apply.)

	Comparison	Unadjusted p-value
A.	A vs B	0.7353
B.	A vs C	0.1872
C.	A vs D	0.2941
D.	B vs C	0.0365
E.	B vs D	0.3677
F.	C vs D	0.0034

For question 75 – 77:

The following example demonstrates the benefit of multiple linear regression. Consider the left plot. It is a scatter plot of `AGE` and `Y`. The linear regression line is plotted and there appears to be positive association between the two variables. Now consider the center and right plot. In these plots, we consider the relationship of `AGE` and `Y` for each gender. Notice that the relationship is between `Y` and `AGE` depends on gender. For males, the association appears to be positive. For females, the association appears to be negative. A multiple linear regression model allows us to model the relationship of `AGE` and `GENDER` with `Y`.



The following computer output provides the estimates of the multiple linear regression model.

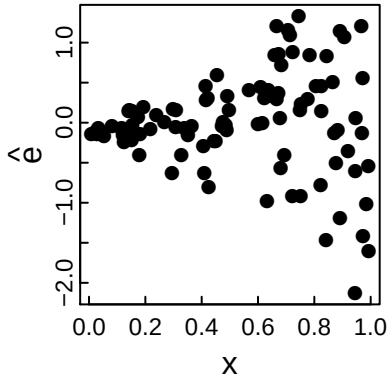
```
                    EST            95% CI
Intercept for F  143.87      ( 124.83,  162.92 )
Intercept for M   51.44      (  34.38,   68.50 )
Slope for F       -0.35      (  -0.70,    0.01 )
Slope for M        1.03      (   0.70,    1.35 )
---
n = 100, k = 4
residual sd = 29.75, R-Squared = 0.94
```

75. *(2 pts)* Perform a test of linear association between AGE and Y for MALES at an $\alpha = 0.05$ test level. What do you conclude?
    - `A.` The data suggests AGE and Y are linearly associated for MALES.
    - `B.` The data suggests AGE and Y are NOT linearly associated for MALES.
    - `C.` Neither. Cannot perform a linear association test with a regression model.

76. *(2 pts)* Perform a test of linear association between AGE and Y for FEMALES at an $\alpha = 0.05$ test level. What do you conclude?
    - `A.` The data suggests AGE and Y are linearly associated for FEMALES.
    - `B.` The data suggests AGE and Y are NOT linearly associated for FEMALES.
    - `C.` Neither. Cannot perform a linear association test with a regression model.

77. *(2 pts)* Suppose you know the age of an additional female is 20. Use the linear regression model to estimate Y for this additional individual.

78. *(2 pts)* The following type of plot is best used for which purpose?



A. Check the normality of the outliers
B. Check the normality of the residuals
C. Investigate heteroskedacity
D. None of the above

79. *(2 pts)* The items that may beneficially affect the validity of a simple linear regression are _____. (Mark all that apply.)

A. A strong correlation between X and Y
B. Outliers
C. Heteroskedasticity
D. The errors are normally distributed
E. None of the above

■ Read the abstract and answer the questions that follow.

The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

A Two-Year Randomized Trial of Obesity Treatment in Primary Care Practice

Thomas A. Wadden, Ph.D., Sheri Volger, R.D., M.S., David B. Sarwer, Ph.D., Marion L. Vetter, M.D., R.D., Adam G. Tsai, M.D., Robert I. Berkowitz, M.D., Shiriki Kumanyika, Ph.D., M.P.H., Kathryn H. Schmitz, Ph.D., M.P.H., Lisa K. Diewald, R.D., M.S., Ronald Barg, M.D., Jesse Chittams, M.S., and Reneé H. Moore, Ph.D.

ABSTRACT

BACKGROUND

Calls for primary care providers (PCPs) to offer obese patients behavioral weight-loss counseling have not been accompanied by adequate guidance on how such care could be delivered. This randomized trial compared weight loss during a 2-year period in response to three lifestyle interventions, all delivered by PCPs in collaboration with auxiliary health professionals (lifestyle coaches) in their practices.

METHODS

We randomly assigned 390 obese adults in six primary care practices to one of three types of intervention: usual care, consisting of quarterly PCP visits that included education about weight management; brief lifestyle counseling, consisting of quarterly PCP visits combined with brief monthly sessions with lifestyle coaches who instructed participants about behavioral weight control; or enhanced brief lifestyle counseling, which provided the same care as described for the previous intervention but included meal replacements or weight-loss medication (orlistat or sibutramine), chosen by the participants in consultation with the PCPs, to potentially increase weight loss.

RESULTS

Of the 390 participants, 86% completed the 2-year trial, at which time, the mean (±SE) weight loss with usual care, brief lifestyle counseling, and enhanced brief lifestyle counseling was 1.7±0.7, 2.9±0.7, and 4.6±0.7 kg, respectively. Initial weight decreased at least 5% in 21.5%, 26.0%, and 34.9% of the participants in the three groups, respectively. Enhanced lifestyle counseling was superior to usual care on both these measures of success ($P=0.003$ and $P=0.02$, respectively), with no other significant differences among the groups. The benefits of enhanced lifestyle counseling remained even after participants given sibutramine were excluded from the analyses. There were no significant differences between the intervention groups in the occurrence of serious adverse events.

CONCLUSIONS

Enhanced weight-loss counseling helps about one third of obese patients achieve long-term, clinically meaningful weight loss. (Funded by the National Heart, Lung, and Blood Institute; POWER-UP ClinicalTrials.gov number, NCT00826774.)

From the Departments of Psychiatry (T.A.W., S.V., D.B.S., M.L.V., R.I.B., L.K.D., R.H.M.), Surgery (D.B.S.), and Medicine (M.L.V.) and the Center for Clinical Epidemiology and Biostatistics (S.K., K.H.S., J.C., R.H.M.), the Perelman School of Medicine at the University of Pennsylvania; the Department of Child and Adolescent Psychiatry, Children's Hospital of Philadelphia (R.I.B.); and Penn Medicine, Clinical Care Associates (R.B.) — all in Philadelphia; and the Division of General Internal Medicine and Center for Human Nutrition, University of Colorado, Aurora (A.G.T.). Address reprint requests to Dr. Wadden at the University of Pennsylvania, 3535 Market St., Suite 3029, Philadelphia, PA 19104, or at wadden@mail.med.upenn.edu.

This article (10.1056/NEJMoa1109220) was published on November 14, 2011, at NEJM.org.

N Engl J Med 2011;365:1969-79.

80. *(1 pt)* What is the study design?

    A. Observational survey
    B. Randomized, control trial

81. *(2 pts)* How many intervention groups does the study have?

82. *(2 pts)* In the results section, the authors report two types of outcome variables. Which did they report? (Mark all that apply.)

    A. Change in Body-Fat Percentage (BFP)
    B. Change in Body-Mass Index (BMI)
    C. Change in Weight (kg)
    D. Change in Obesity Status (Yes/No)
    E. Weight Loss of 5% or more (Yes/No)

83. *(3 pts)* What is the central contradiction of public health ethics?

84. *(2 pts)* Enjoy 2 free points.