**Multiple Choice [12 pts]-- choose the ONE BEST answer for each question by writing the corresponding letter in the blank to the left of the question.**

1. What is the name of the variable that we are interested in predicting or explaining?

**D B** **A.** Continuous **B.** Discrete **C.** Explanatory **D.** Response **E.** Sampling

1. Which term is defined as “the change in value of the response variable for a unit change in the explanatory variable?”

**D E** **A.** Correlation **B.** Y-Intercept **C.** RSS **D.** Slope **E.** r2

1. What is the vertical difference between an observed and predicted value of the response variable?

**D B** **A.** Explanatory **B.** Frequency **C.** Predicted **D.** Residual **E.** Response

1. Which of the following is NOT a possible value for the coefficient of determination?

**A E**  **A.** -0.6 **B.** 0.34789 **C.** 0.1 **D.** 0 **E.** 1

1. What is it called when you are asked to make a prediction outside the domain of the explanatory variable?

**C A** **A.** Residual **B.** Response **C.** Extrapolation **D.** Homoscedastic **E.** Heteroscedastic

6. Which word best describes the situation where the best-fit line goes through the middle of the points on a scatterplot but the points are unequally scattered around that line?

**E C** **A.** Linear **B.** Non-linear **C.** RSS **D.** Homoscedastic **E.** Heteroscedastic

-

7. What type of study is it if the researcher sends a survey to the entire population, but only some of the individuals return it?

**E E** **A.** Convenience **B.** Inference **C.** Simple Random **D.** Regression **E.** Voluntary Response

8. What type of study is it if every individual has the same chance of being selected for the sample?

**C A** **A.** Convenience **B.** Inference **C.** Simple Random **D.** Regression **E.** Voluntary Response

9. What is the number of combinations of all factors in a multi-factor experiment?

**E C** **A.** Response **B.** Levels **C.** Replicates **D.** Placebo **E.** Treatments

-

10. What is the measure of dispersion among statistics from all possible samples?

**D B** **A.** Response **B.** Mean **C.** Replicates **D.** Standard Error **E.** Standard Deviation

11. What is the measure of dispersion on a symmetric population distribution?

**E E** **A.** Response **B.** Mean **C.** Replicates **D.** Standard Error **E.** Standard Deviation

12. What is the measure of sampling variability?

**D B** **A.** Response **B.** Mean **C.** Replicates **D.** Standard Error **E.** Standard Deviation

**library(NCStats)**

**distrib(x,mean=##,sd=##,lower.tail=XXXXX,type=”X”)**

where **x** is replaced with the value of the quantitative variable or the area

**mean=##** has ## replaced by the value of the mean

**sd=##** has ## replaced by the value of the standard deviation or standard error

**lower.tail=XXXXX** has XXXXX replaced with TRUE (default) for a “left-of” and FALSE for a “right-of” calculation

**type=”X”** has X replaced with p (default) for a forward and q for a reverse question

**sqrt(x)**

**Answer the following questions in the space provided. Show your work where appropriate.**

1. **[11 pts]** Vasey *et al.* (2012; J Anxiety Disord) examined subjects that had a self-described fear of spiders. Each subject was shown a spider and asked to estimate its size. The subject’s estimate of the size of the spider was divided by the actual size of the spider, which they called “spider size discrepancy” (values <1 are underestimates, values >1 are overestimates of the size of the spider). Additionally, the researchers recorded the subject’s level of distress upon seeing the spider on a scale from 0-100 (higher values mean more distress). The researchers were interested in determining if the subject’s distress level could explain their spider size discrepancy estimate. Use their results in Figure 1 or 2 (as appropriate) to answer the questions below the figures.



**Figure 1.** Linear regression of a subject’s distress on the **Figure 2.** Linear regression of a subject’s discrepancy

discrepancy in the estimated size of the spider. In size of the spider on their distress level.

1. In terms of the variables of this problem, what is the equation of the best-fit line?

**Discrepancy = 0.0036\*Distress+0.778**

1. In terms of the variables of this problem, interpret the value of the slope?

**For each one unit increase in the subject’s level of distress, the size discrepancy increases by 0.0036 units, ON AVERAGE**

1. What is the predicted spider size discrepancy if the subject’s distress level is 80?

Discrepancy = 0.0036\*80+0.778 = **1.066**

1. What is the residual if the spider size discrepancy is 0.7 and the subject’s distress level is 30?

Discrepancy = 0.0036\*30+0.778 = 0.886 Residual = 0.7-0.886 = **-0.186**

1. How much would one expect a subject’s spider size discrepancy to change if they could reduce their distress level by 20 points?

-20\*slope = -20\*0.0036 = **-0.072**

1. What percentage of the variability in spider size discrepancy is explained by the subject’s distress level?

**r2 = 0.126**

1. **[8 pts]** Researchers observed groups of dolphins off the coast of Iceland near Keavik in 1998. The researchers recorded the time of the day (Morning, Noon, Afternoon, and Evening) and the main activity of the group (travelling, feeding, or socializing). The number of dolphin groups observed by each time of day and activity is shown in Table 1. Use these results to answer the questions below the table.

Table 1. Frequency of Dolphin groups by time of day and type of activity.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Type of Activity** | | |  |
| **Time of Day** | **Traveling** | **Feeding** | **Socializing** |  |
| **Morning** | 6 | 28 | 38 | 72 |
| **Noon** | 6 | 4 | 5 | 15 |
| **Afternoon** | 14 | 0 | 9 | 23 |
| **Evening** | 13 | 56 | 10 | 79 |
|  | 39 | 88 | 62 | 189 |

1. What percentage of all dolphin groups were observed in the evening?

79/189\*100 = **41.8%**

1. What percentage of dolphin groups observed in the evening were feeding?

56/79\*100 = **70.9%**

1. What percentage of socializing dolphin groups were observed in the morning?

38/62\*100 = **61.3%**

1. What percentage of dolphin groups observed in the morning were feeding?

28/72\*100 = **38.9%**

1. **[7 pts]** Maret and Collins (1996) set up an experiment to test the effects of food ration (low, medium, and high densities of brine shrimp per day) and the density of a competitor (0, 1, or 2 frog tadpoles present) on the growth of larval salamanders. A total of 45 identical aquaria that could be maintained in a controlled laboratory environment were available for the experiment. Use this information to answer the following questions.

a. What is the response variable? \_**growth of larval salamanders**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

b. What is/are the factor(s)? \_**food ration and density of brine shrimp**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

c. What is/are the number of levels? \_**3** for food ration **and 3** for density of brine shrimp­­­­­­­­\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

d. What is the number of treatments? \_**9** (3x3)\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

e. What is the number of replicates per treatment? \_**5** (45/9)\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Use distrib() in RStudio to produce the result(s) needed to answer the next question. Write your answers with complete sentences with the code used to produce the result below your sentence.**

1. **[8 pts]** Assume that a wholesale distributor of fertilizer products knows that the weekly demand for 5-Nitrogen, 5-Phosphorous, 6-Potassium fertilizer is slightly right-skewed with a mean of 12 tons and a standard deviation of 3.6 tons. Use this information to answer the questions below. [Note: R suggestions are on the first page.]
2. What is the probability that the demand for one week will exceed 14 tons?

**Can’t answer, population is not normally distributed**

1. What is the probability that the mean demand for a sample of 20 weeks will exceed 14 tons?

> ( distrib(14,mean=12,sd=3.6/sqrt(20),lower.tail=FALSE) )

[1] **0.006486311**

1. What is the probability that the mean demand for a sample of 40 weeks will be between 11.5 and 13 tons?

> ( ab <- distrib(13,mean=12,sd=3.6/sqrt(40)) )

[1] 0.9605258

> ( a <- distrib(11.5,mean=12,sd=3.6/sqrt(40)) )

[1] 0.1898605

> ab-a

[1] **0.7706653**

1. What is the most common 90% of sample means from n=40 weeks?

> ( distrib(0.05,mean=12,sd=3.6/sqrt(40),type="q") )

[1] **11.06373**

> ( distrib(0.95,mean=12,sd=3.6/sqrt(40),type="q") )

[1] **12.93627**

**Short (Paragraph) Answers -- Answer THREE of the following questions with complete sentences on a separate sheet of paper. Circle the questions below that you have chosen to answer and make sure to clearly label your answers on the separate sheet. Each question is worth 3 points.**

1. What are the two major goals of regression? Give a specific example to illustrate each goal.

Predict and Explain Variability

1. What are the two major assumptions of regression? Draw plots that illustrate the meeting and violating of these assumptions.

Linearity (obvious curve), Homoscedasticity (funnel shape), Outlier

Meeting assumptions (complete random scatter, no pattern)

1. What are the three major principles of experimental design and why is each important?

Control – isolate effects of factor(s) on response

Randomization – remove bias, allow laws of probability to be used

Replication – factor out effect of single individuals (a form of control)

1. Of what value are observational studies? Give specific examples.

Cannot conduct an experiment (cost, size, no possible replicates, ethical issues with experiment)

Making observation prior to experiment (first step in scientific method)

Validating results of experiment in real world

1. Completely describe all differences between a population and a sampling distribution.

Population – for individuals, uses SD for natural variability, specifics must be given or assumed

Sampling – for statistics, uses SE for sampling variability, specifics from CLT