

Part I: Answer three of the following four questions. If you answer more than three I will grade only the first three. Five points each.

1) Define what is meant by the p-value (or observed significance level) of a test. **The p-value is the probability of observing a statistic as extreme as the one observed, or more extreme, if the null hypothesis is true. -or- The p-value is the smallest α -level at which the null hypothesis would be rejected.**

2) The histogram at the right is **skewed left** We would expect that its mean would be **smaller than** its median.

3) A fair coin (probability of a head on one flip = 0.5) is flipped 10 times. What is the probability of observing exactly

$$5 \text{ heads? } \binom{10}{5} 0.5^5 (1 - 0.5)^5 = \frac{10!}{5!5!} 0.5^{10} = 252(0.000976562) \approx 0.2461$$

4) A fair coin (probability of a head on one flip = 0.5) is flipped 1,000 times. Using the normal approximation to the binomial, approximately what is the probability of observing at least 512 heads?

$$Z = \frac{(X - np)}{\sqrt{np(1 - p)}} = \frac{(512 - 0.5) - 1000(0.5)}{\sqrt{1000(0.5)(1 - 0.5)}} = \frac{11.5}{15.811} \approx 0.73$$

We need to subtract the 0.5 for the continuity correction. The answer we want is $P(Z > 0.73) = 0.5 - 0.2673 = 0.2327$

Part II: Answer every part of the next three problems. Read each problem carefully, and show your work for full credit. Twenty points each.

1A) State the appropriate null and alternate hypothesis to test if a high score on the exam seems to indicate mild to moderate traumatic brain injury. Identify any parameters you use in stating the hypotheses.

$H_0: \mu = 40$ vs. $H_A: \mu > 40$ where μ is the average test score of all patients with mild to moderate traumatic brain injuries

B) Test the hypothesis in part a. Report your conclusion at $\alpha = 0.05$ (e.g. Do we accept or reject H_0 ? Is a high test score associated with traumatic brain injury, or not?)

$$t_{df=n-1=22} = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{48.43 - 40}{20.76/\sqrt{23}} = \frac{8.43}{4.32} = 1.95 \quad \text{compare this to 1.717 and reject } H_0. \quad \text{The injured do seem to score higher.}$$

C) Construct a 95% confidence interval for the mean test score of patients who suffered mild to moderate traumatic brain injury.

$$\bar{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}} = 48.43 \pm 2.074 \frac{20.76}{\sqrt{23}} = 48.43 \pm 8.98 = (39.45, 57.41)$$

D) What assumptions need to be satisfied in order to trust the results you obtained in parts B and C?

The patients need to be a random sample and their test scores need to follow a normal distribution.

2A) Note that four values have been deleted from the ANOVA table. What values should they have?

Summary of Fit					
Mean of Response		81.0714	R-Square	0.5880	
Root MSE		8.7867	Adj R-Sq	0.5537	

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Stat	Pr > F
Model	1	1322.4642	1322.4642	17.1289	0.0014
Error	12	926.4644	77.2054		
C Total	13	2248.9286			

B) Assuming the assumptions for predicting this linear regression are met, perform the test of hypotheses for testing whether batting average predicts wins. What is the p-value, and what is your conclusion $\alpha=0.05$?

The p-value is 0.0014. We reject the null hypothesis and conclude that batting average does predict wins.

C) Assuming the assumptions for this simple linear regression are met, identify which of the following statements are true or false.

- F 58.8% of a team's wins are explained by batting average.
- T 58.8% of the variation in the number of wins a team has is explained by the batting average.
- F The correlation coefficient for this regression is 0.588.
- T The coefficient of determination for this regressions is 0.588.

D) Assuming the assumptions for this simple linear regression are met, give the estimated range that the model says 95% of the teams batting 0.300 should have win totals between. **From the prediction interval, 85.9738 to 136.8921.**

E) Why can't we trust the interval in D? **A 0.300 batting average is out of the range of our data (we would be extrapolating.)**

	Male Viewer	Female Viewer	
Identified Product	95	41	136
Could Not Identify Product	55	109	164
	150	150	300

A) Would this data be analyzed by using a test of independence, **a test of homogeneity**, or a goodness of fit test?

B) Write out the tables of expected values for conducting this test.

	Male Viewer	Female Viewer
Identified Product	(136)(150)/300=68	(136)(150)/300=68
Could Not Identify Product	(150)(164)/300=82	(150)(164)/300=82

C) Give the formula for X^2 for this problem (plugging the values in, but not needing to simplify).

$$X^2 = \frac{(95 - 68)^2}{68} + \frac{(41 - 68)^2}{68} + \frac{(55 - 82)^2}{82} + \frac{(109 - 82)^2}{82}$$

D) What is the rejection region (critical region) for conducting this test at $\alpha=0.05$? **df=1, so Reject if $X^2 \geq 3.84146$.**

E) Why is, or why isn't, the sample size of this experiment large enough for performing this hypothesis test?

It is large enough because all of the expected values are at least 5.