Amherst Statistics Final Solutions

Simon Rubinstein-Salzedo

May 30, 2013

I make no guarantees about the correctness of these solutions; use at your own risk!

Part I:

1. D
2. C
3. B
4. Unclear, but probably A (χ^2 should be for proportions, although we can turn this into a proportion)
5. B
6. E
7. C
8. C
9. C
10. D (not part of the class)

Part II:

1. (a) The point estimate is $\widehat{p} = \frac{46}{452} = .1018$. The standard error is $\sqrt{(.1018)(.8982)/452} = .0142$. Hence the 95% confidence interval is $.1018 \pm 1.96 \times .0142$.

(b) If we collected many similar samples and constructed confidence intervals for all of them, then the true proportion would lie in the interval 95% of the time.

2. (a) Analysis: χ^2

   Null: Marijuana usage and religiosity are independent.

   Alternative: Marijuana usage and religiosity are not independent.

(b)

ReligImp \ Use Marijuana	Yes	No	Total
Fairly important	102 (95.77)	120 (126.23)	222
Not important	68 (55.22)	60 (72.78)	128
Very important	25 (44.00)	77 (58.00)	102
Total	195	257	452

   (c) $(R-1)(C-1) = 2$.

   (d) We reject the null hypothesis that marijuana usage and religiosity are independent.

   (e) $(25-44)^2/44 = 8.20$. This is certainly a large value for a single cell, so we suspect that very religious people might be less like to use marijuana than average.

3. (a) One-sided difference of means test

   (b) H_0: There is no difference between the average days per month that Greek and non-Greek students drink.

   H_A: Greek students drink more days per month than do non-Greek students.

   (c) p-value: 0 (or very close to it); Conclusion: we believe that Greek students drink more days per month than do non-Greek students.

   (d) We may have falsely rejected the null hypothesis, which is a Type I error. This would happen if it were not true that Greek students drink more days per month than do non-Greek students.

4. (a) H_0: The mean study time is the same in all groups.

   H_A: The mean study time is not the same in all groups.

   (b) Observations are independent within and between groups, data within each group are nearly normal, variability across groups is about equal.

(c) $F = 572/93 = 6.15$, $p = 0.002326$.

(d) We do not believe that the mean study time is the same in all groups.

(e) We can try, using the Bonferroni correction.

(f) (Not relevant for us.)

5. (a) $R^2 = .516$. This measures how strong the linear relationship is.

(b) $\widehat{y} = 2.48541 + .83553x$, where x is the number of party days and y is the number of alcohol days.

(c) Residual is $8 - (2.48541 + .83553 \times 8) = -1.16965$.

(d) H_0: The true slope is zero.

H_A: The true slope is nonzero.

Hypotheses are okay. Test statistic is 21.901, and p-value is under 2×10^{-16}. We therefore reject H_0.

(e) $.83553 \pm 1.96 \times .03815$. 95% of the time when we collect data, the true slope will lie in the interval we construct.

6. (a) For high alcohol days: Back 27%, Middle 57%, Front 16%. For low alcohol days: Back 15%, Middle 57%, Front 28%.

(b) I don't really feel like running a χ^2 test here, but it appears that students with low alcohol days sit in the front more often than do students with high alcohol days, and the reverse is true for the back.

(c) Point estimate is $.27 - .15 = .12$. Standard error is

$$SE = \sqrt{\frac{.27 \times .73}{218} + \frac{.15 \times .85}{234}} = .038.$$

The confidence interval is $.12 \pm 1.96 \times .038$.

7. (a) You can pair a person's right hand span with the same person's left hand span.

(b) H_0: The average difference between left hand span and right hand span is 0.

H_A: The average difference between left hand span and right hand span is greater than 0.

(c) Sample size at least 30, independence, not too much skew.

(d) At the 95% confidence level, we fail to reject H_0, since $p > 0.05$.

(e) p-value is the probability that we get a result at least as favorable to H_A by chance, assuming that H_0 is true.

8. (a) Right skewed

(b) The point at 2000 is clearly an outlier.

(c) Median and IQR, because they are not sensitive to outliers.

(d)
$$Z = \frac{120 - 110}{32/\sqrt{100}} = 3.125.$$

The probability of a Z-score above 3.125 is 0.0009.