Stat 252 Exam 1 January 27, 2010

Note: The order of questions on your exam may differ. Also, check for whether you have the alternate version of some questions.

Reese's Pieces question:

*. (7 pts) Recall that we collected data in class on colors of Reese's Pieces candies. We used the data to test the hypothesis that 50% of all Reese's Pieces are orange, 25% are brown, and 25% are yellow.

a) (2 pts) I mentioned to a colleague that we did not reject this hypothesis at the $\alpha = .10$ level. Based only on this information (and not on the actual data), can my colleague decide whether or not we would have rejected the null hypothesis at the $\alpha = .01$ level? Explain briefly.

Yes. Because we failed to reject the hypothesis at the $\alpha = .10$ level, we know that p-value > .10. So, the p-value is also > .01 and so we would also fail to reject the hypothesis at the $\alpha = .01$ level.

Alternate version:

a) I mentioned to a colleague that we did not reject this hypothesis at the $\alpha = .05$ level. Based only on this information (and not on the actual data), can my colleague decide whether or not we would have rejected the null hypothesis at the $\alpha = .10$ level? Explain briefly.

No. Because we failed to reject the hypothesis at the $\alpha = .05$ level, we know that p-value > .05. But it's possible that p-value > .10 and it's also possible that .01 < p-value < .10, so my colleague cannot determine whether we would have rejected the null hypothesis at the $\alpha = .10$ level.

b) (1 pt) Because we did not reject this hypothesis, is it reasonable to conclude that if we had tested a different hypothesis about the proportional breakdown of colors (such as: 40% orange, 30% brown, 30% yellow), we would have rejected that hypothesis? (Circle your answer; do not bother to explain.) Yes No

It's certainly possible that there would be many null hypotheses that would not be rejected.

c) (2 pts) My colleague asked if we might have committed a Type II error. Explain what a Type II error means in this context.

Type II error is to fail to reject the null hypothesis when the null hypothesis is actually false. In this case, type II error is to fail to reject the 50-25-25 hypothesis for the proportions of orange-brown-yellow when in fact the color proportions are not 50-25-25.

d) (2 pts) Suggest one way that we could have altered this study in order to obtain a more powerful test. (Do not bother to explain.)

Using a larger sample (more candies) would have produced a more powerful test, as would using a larger significance level.

Organ donor question:

*. (7 pts) I recently read an article in *Science* magazine that described an experiment concerning people's willingness to be organ donors. All of the subjects in the experiment were told that to imagine that they have moved to a new state and have applied for a driver's license, and they must make a decision about whether to become an organ donor. Some of the subjects were randomly assigned to be told that the default option is not to be a donor, and the rest of the subjects were told that the default option is to be a donor. All subjects were then given the choice of whether or not to become an organ donor. The researchers suspected that a higher proportion of people are willing to be donors when the default option is to be a donor.

a) (2 pts) Identify the explanatory and response variables in this study.

Explanatory:

The explanatory variable is the default option presented (either to be a donor or not)

Response:

The response variable is the person's decision about whether or not to become an organ donor.

b) (2 pts) State (in symbols) the appropriate null and alternative hypotheses (in symbols) for testing the researchers' suspicion.

H_0: $\pi_{\text{default = donor}} = \pi_{\text{default = not donor}}$ H_a: $\pi_{\text{default = donor}} > \pi_{\text{default = not donor}}$

c) (3 pts) The article reported that 42% of the subjects in the "default is not to be a donor" group decided to become a donor, compared to 82% in the "default is to be a donor" group. What further information would you need in order to conduct the test (i.e., to calculate a test statistic and p-value)?

You need to know the number of people (sample size) in each group.

Close friends question:

*. (9 pts) One of the questions asked of a random sample of 1467 adult Americans in the 2004 General Social Survey (GSS) was: "From time to time, most people discuss important matters with other people. Looking back over the last six months - who are the people with whom you discussed matters important to you? Just tell me their first names or initials."

The interviewer then recorded how many names or initials the respondent mentioned. For convenience, we will refer to this as the person's number of "close friends." Suppose that we want to examine whether men and women differ on average with regard to how many close friends they have.

a) (2 pt) The null hypothesis can be written as H_0: $\mu_{men} = \mu_{women}$. Explain what the symbol μ_{men} represents here.

This symbol represents the mean number of close friends in the population of all American men.

b) (1 pt) Write the appropriate alternative hypothesis in symbols.

H_a: $\mu_{men} \neq \mu_{women}$

The sample data are reported here:

Number of close friends	0	1	2	3	4	5	6	Total
Number of respondents (male)	196	135	108	100	42	40	33	654
Number of respondents (female)	201	146	155	132	86	56	37	813

c) (2 pts) One technical condition for the two-sample *t*-test is that the data arise from either random sampling or random assignment. Does this study involve both of these, only one of them, or neither? (If your answer is "only one," be sure to specify which.)

This study makes use of random sampling only. It's not possible to randomly assign gender to people.

d) (2 pts) The other technical condition is that the sample sizes are large or the data follow a normal distribution. Are both of these satisfied, or only one, or neither? Justify your answer.

The sample size condition is met, because both sample sizes (654, 813) are much larger than 30. But the table reveals that both distributions (of number of close friends) are skewed to the right, so the normality part of the condition is not satisfied.

e) (2 pts) The test statistic turns out to be $t = -2.45$, and the *p*-value is .014. Based on this information alone, what can you say about a 95% confidence interval for $\mu_{men} - \mu_{women}$? Explain.

Because the p-value is less than .05, we know that μ_{men} differs significantly from μ_{women} at the .05 level. So, the 95% confidence interval for $\mu_{men} - \mu_{women}$ will not include 0. Because the test statistic is negative, we know that this interval will include only negative values, indicating that $\mu_{men} < \mu_{women}$.

Enzyme therapy question:

*. (6 pts) In October of 2009, actress Suzanne Somers published a book titled *Knockout: Interviews with Doctors Who are Curing Cancer and How to Prevent Getting It*. In this book she advocates alternative methods of treating cancer. One doctor that she especially praises is

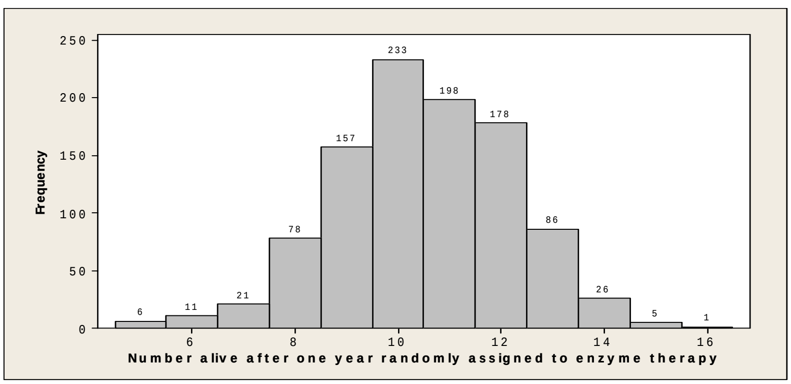
Dr. Nicholas Gonzalez, who uses an enzyme therapy technique to treat pancreatic cancer. Dr. Gonzalez's method was tested in a study published in the *Journal of Clinical Oncology* on August 17, 2009. This study involved 55 patients who had inoperable pancreatic cancer. Each patient was given the choice of chemotherapy or enzyme therapy. It turned out that 23 chose chemotherapy and 32 opted for enzyme therapy. Then results are summarized here:

	Enzyme therapy	Chemotherapy	Total
Still alive at end of year	5	13	18
Died during year	27	10	37
Total	32	23	55

a) (1 pt) Is this an observational study or an experiment? (Circle your answer; do not bother to explain.) Observational study Experiment

Because subjects choose for themselves which explanatory variable group (enzyme therapy or chemotherapy) to be in, this is an observational study. It's also reasonable to consider this an experiment, because a therapy is imposed on the subjects, although it is not a *randomized* experiment.

The following histogram displays simulation results of randomly assigning these 55 patients into two groups, assuming that there is no difference in effectiveness between enzyme therapy and chemotherapy:



b) (2 pts) Use this histogram to determine the approximate p-value for testing whether enzyme therapy is less effective than chemotherapy.

The approximate p-value is the fraction of these 1000 repetitions that produced 5 or fewer successes in the enzyme therapy group (because that's how the actual data from the study turned out). This fraction is 6 / 1000 = .006.

c) (3 pts) Summarize the conclusion that you would reach from this study and this analysis, addressing issues of significance and causation.

The very small p-value indicates that enzyme therapy has a significantly lower proportion of patients who live for a year, as compared to the chemotherapy group. But because patients chose their treatment group for themselves, we cannot draw a cause-and-effect conclusion between the enzyme therapy and the lower survival proportion.

Tipping question:

*. (12 pts) Can telling a joke affect whether or not a customer in a coffee bar leaves a tip for the waiter? A recent study investigated this question at a coffee bar at a famous resort on the west coast of France (Gueguen, 2002). The waiter randomly assigned coffee-ordering customers into one of three groups: one group received a card telling a joke with the bill, another group received a card containing an advertisement for a local restaurant, and a third group received no card at all. Results are summarized in the following 2×3table:

	Joke Card	Advertisement Card	No Card	Total
Left a Tip	30	14	16	60
Did Not Leave a Tip	42	60	49	151
Total	72	74	65	211

Analysis of these data produced the following Minitab output (notice that some of the output has been removed):

```
              Joke Card  Advert Card  No Card  Total
Left tip             30           14       16     60
                  xxxxx        21.04    18.48
                  4.432        2.357    0.334

No tip               42           60       49    151
                  51.53        52.96    46.52
                  1.761        0.937    0.133

Total                72           74       65    211

Chi-Sq = 9.953, DF = x, P-Value = xxxxx
```

a) (2 pts) Determine the expected count for the (left tip, joke card) cell. Show how you determine this value, without using any of the expected counts provided in the output.

$60 \times 72 / 211 \approx 20.47$

b) (2 pts) Report the degrees of freedom, and determine the p-value, as accurately as possible from the appropriate table.

The df = $(2-1)\times(3-1) = 2$. Looking for a test statistic value of 9.953, Table X reveals that the p-value is between .005 and .01.

c) (2 pts) Do the data suggest that this waiter's likeliness of receiving a tip differs among these three strategies? Justify your answer.

Yes. The very small p-value (less than .01) indicates strong evidence that the three strategies do not all have the same proportion of customers who leave a tip.

d) (2 pts) What advice, if any, would you give the waiter for trying to increase his likeliness of receiving a tip? Explain briefly, based on this study and analysis.

The waiter should leave a joke card. The largest contribution to the chi-square test statistic comes from the (left a tip, joke card) cell of the table, with many more customers than expected leaving a tip when they received a joke card. The worst option would be to leave an advertisement card, because that strategy produced many fewer tips than expected.

e) (2 pts) Is it reasonable to draw a cause/effect conclusion between the type of card and the likeliness of receiving a tip? Explain (very) briefly.

Yes, because the customers were randomly assigned to groups, and because the very small p-value indicates a significant difference among the groups.

f) (2 pts) Is it reasonable to generalize the results of this study to all waiters in all types of restaurants? Explain (very) briefly.

No, because this study was only conducted on one waiter in one particular establishment in a fairly unusual environment (coffee bar at famous resort on coast of France).

Choose a procedure question:

*. (9 pts, 3 pts each) Some of the statistical inference techniques that we have studied so far include:

A. One-sample z-procedures for a proportion
B. One-sample t-procedures for a mean
C. Two-sample z-procedures for comparing proportions
D. Two-sample t-procedures for comparing means
E. Paired-sample t-procedures
F. Chi-square goodness-of-fit procedures
G. Chi-square procedures for two-way tables

For each of the following questions, identify (by letter) the procedure that you would use to investigate that question. Also indicate (either in symbols or in words) the null and alternative hypothesis to be tested in each case.

*) Some critics of the military have claimed that members of the U.S. Armed Forces have an average IQ less than 100. Suppose that you take a random sample of members of the U.S. Armed Forces and measure their IQs, in order to test this claim.

Procedure: B. One-sample t-procedures for a mean

Null hypothesis: $\mu = 100$ Alternative hypothesis: $\mu < 100$

*) Researchers investigated whether drivers tend to take longer to react to a stimulus if they are talking on a cell phone than if they are listening to an audiobook. A sample of 32 drivers participated in a simulated driving device under both conditions (talking on phone, listening to book), and researchers recorded how long it took to react to a stimulus under each condition.

Procedure: E. Paired-sample *t*-procedures

Null hypothesis: $\mu_{diff\ (phone\ -\ book)} = 0$ Alternative hypothesis: $\mu_{diff\ (phone\ -\ book)} > 0$

*) Do the Boston Celtics allow significantly fewer points on average during games in which Kevin Garnett plays as opposed to games in which Kevin Garnett does not play? To investigate this question, you record the number of points allowed by the Celtics in every game played this year, along with whether or not Garnett played in the game.

Procedure: D. Two-sample *t*-procedures for comparing means

Null hypothesis: $\mu_{KG} = \mu_{no\ KG}$ Alternative hypothesis: $\mu_{KG} < \mu_{no\ KG}$

Alternate:

*) Are employee sick days for a particular company equally likely to be used on the five days of the workweek (Monday – Friday)? You take a random sample of 250 employee sick days and record the day of the week for each.

Procedure: F. Chi-square goodness-of-fit procedures

Null hypothesis: $\pi_m = .2, \pi_{tu} = .2, \pi_w = .2, \pi_{th} = .2, \pi_f = .2$ Alternative: not H_0