2013 Autumn STAT 22000 Session 02 Final Exam Solutions

1. [Nutrition] [12 points]

A multimedia program designed to improve dietary behavior among low-income women (those who live on food stamps) was evaluated by taking a random sample from Food Stamp recipients in Durham, North Carolina, and having them watch a 30-minute session in a computer kiosk in the Food Stamp office. One of the outcomes was the score on a nutrition knowledge test taken about 2 months after the program. Here is a summary of the data:

Sample Size	Sample Mean	Sample SD
101	5.08	1.15

(a) [2 points] The test had six multiple-choice items that were scored as correct or incorrect, so the total score was an integer between 0 and 6. Do you think that these data are Normally distributed? Explain why or why not.

Answer: No. The scores take only take value 0, 1, 2, 3, 4, 5, and 6, which cannot be normal.

Alternatively, one can also argue using the 68-95-99.7 rule. If the data is normal, about 16% of the data will be at least 1 SD above the mean, i.e., at least $5.08 + 1.15 = 6.23$, which is impossible because the total score can not go over 6.

The sample appear left-skewed because the distance from the mean 5.08 to the max 6 is smaller than the SD, so the left tail must be longer than the right tail.

(b) [2 points] Is it appropriate to use the one-sample t procedures to analyze these data? Explain why or why not.

Answer: Yes. The t-procedure is robust to some moderate non-normality, even of the data is skewed. Moreover, the sample size 101 is large, in which case the t-procedure can be used even for clearly skewed data. See "Robustness of the t-procedures" on p.417-418 of the textbook.

(c) [5 points] Construct a 95% confidence interval for the mean score in the test of the Food Stamp recipients in Durham, North Carolina.

Answer: As stated in part (b), we can use a one-sample t-interval. The number of degrees of freedom is $101-1=100$. From the t-table, we can see the critical value $t^* = t_{0.025,100} = 1.984$. The interval is

$$\overline{x} \pm t_{0.025,100}\frac{s}{\sqrt{n}} = 5.08 \pm 1.984 \times \frac{1.15}{\sqrt{101}} \approx 5.08 \pm 0.227 = 5.307 \text{ to } 4.853$$

(d) [3 points] Explain to someone who knows no statistics what a 95% confidence interval in part (c) means.

Answer: If we were to repeat the same sampling procedure to get any sample of the same size, in this case, 101, from the same population, and construct confidence interval use the formula,

$$\text{sample mean} \pm 1.984 \times \frac{\text{sample SD}}{\sqrt{101}}$$

many times, about 95% of the constructed interval will contain the population mean score.

2. [Quit Smoking Or Not] [20 points]

A study was done to identify factors affecting physicians' decisions to advise or not to advise patients to stop smoking (Cummings et al. 1987). The study was related to a training program to teach physicians ways to counsel patients to stop smoking and was carried out in a family practice outpatient center in Buffalo, New York. The study population consisted of the cigarette-smoking patients of residents in family medicine seen in the center between February and May 1984.

We first consider whether certain patient characteristics are related to being advised or not being advised. The following table shows a breakdown by gender of the patient:

	Advised	Not Advised
Male	48	47
Female	80	136

(a) [4 points] What proportion of the males were advised to quit and what proportion of the females were advised? What are the standard errors of these proportions? (As the sample size is fairly large, do not use the Wilson plus four estimate.)

Answer: For males, the proportion is $\widehat{p}_m = \dfrac{48}{48+47} = \dfrac{48}{95} \approx 0.5053$. The standard error is

$$SE_m = \sqrt{\frac{\widehat{p}_m(1-\widehat{p}_m)}{n_m}} = \sqrt{\frac{0.5053 \times 0.4947}{95}} \approx 0.0513.$$

For males, the proportion is $\widehat{p}_f = \dfrac{80}{80+136} = \dfrac{80}{216} \approx 0.3704$. The standard error is

$$SE_f = \sqrt{\frac{\widehat{p}_f(1-\widehat{p}_f)}{n_f}} = \sqrt{\frac{0.3704 \times 0.6296}{216}} \approx 0.0329.$$

(b) [6 points] Test whether the proportion of the males were advised to quit equals to the proportion of the females. State the null and alternative hypotheses, give an appropriate test statistic, and report the P-value. (Again, do not use the Wilson plus four estimate.)

Answer: Let p_m be the proportion of male patients being advised to quit smoking, and p_f be the corresponding proportion for female patients. The hypotheses are

$$H_0: p_m = p_f \quad \text{v.s.} \quad H_a: p_m \neq p_f.$$

Under H_0, the estimate of the common p is the pooled sample proportion,

$$\widehat{p} = \frac{48+80}{95+216} = \frac{128}{311} \approx 0.4116.$$

The standard error of the difference under H_0 is then

$$\mathrm{SE} = \sqrt{\widehat{p}(1-\widehat{p})\left(\frac{1}{n_m} + \frac{1}{n_f}\right)} = \sqrt{0.4116 \times 0.5884 \left(\frac{1}{95} + \frac{1}{216}\right)} \approx 0.0606$$

The test statistic is

$$z = \frac{\widehat{p}_m - \widehat{p}_f}{\mathrm{SE}} \approx \frac{0.5053 - 0.3704}{0.0606} = \frac{0.1349}{0.0606} \approx 2.23$$

The 2-sided P-value is $2P(Z > 2.23) = 2P(Z < -2.23) = 2 \times 0.0129 = 0.0258$.

Do physicians of different ages have different tendency to advise patients quit? This table below gives a breakdown by age of physician.

Age of Physician	Advised	Not Advised	Row Total
less than 30	88	128	216
30 to 39	28	37	65
40 and over	12	18	30
Column Total	128	183	311

The usual test statistic for testing the null hypothesis that "the proportions of patients advised to quit smoking are the same regardless of the age of the physician" has been be calculated to be:

$$\text{Test-statistic} = 0.131$$

(c) [2 points] What is the approximate distribution of the test-statistic above under the null hypothesis?

Answer: χ^2-distribution with $df = (r-1)(c-1) = (3-1)(2-1) = 2$.

(d) [6 points] Show how the test-statistic 0.131 is calculated.

Answer: The expected counts are (row total)$\times$(column total)/(overall total) as follows

Age of Physician	Advised	Not Advised	Row Total
less than 30	$\frac{128 \times 216}{311} \approx 88.90$	$\frac{183 \times 216}{311} \approx 127.10$	216
30 to 39	$\frac{128 \times 65}{311} \approx 26.75$	$\frac{183 \times 65}{311} \approx 38.25$	65
40 and over	$\frac{128 \times 30}{311} \approx 12.35$	$\frac{183 \times 30}{311} \approx 17.65$	30
Column Total	128	183	311

The test statistic is the chi-square statistic

$$\begin{aligned}\chi^2 &= \sum_{\text{all cells}} \frac{(\text{Observed count} - \text{Expected count})^2}{\text{Expected count}} \\ &= \frac{(88-88.90)^2}{88.90} + \frac{(128-127.10)^2}{127.10} + \frac{(28-26.75)^2}{26.75} + \frac{(37-38.25)^2}{38.25} + \frac{(12-12.35)^2}{12.35} + \frac{(18-17.65)^2}{17.65} \\ &= 0.131\end{aligned}$$

(e) [2 points] Using the observed test statistic (= 0.131) reported above, answer the question "Do physicians of different ages have different tendency to advise patients quit?" Use significance level $\alpha = 10\%$.

Answer: No. Look at the row in the chi-square table for df = 2. The χ^2-statistic 0.131 is less than the smallest value 2.77 in the row, which means the P-value is at least 0.25. So H_0 is not rejected. There is no evidence that physicians of different age advise patients differently in terms of smoking.

3. [Heat Enduring Glass] [18 points]

A firm producing plate glass has developed a new process meant to allow glass for fireplaces to rise to a higher temperature before breaking. To test the process, plates of glass are drawn randomly from a production run. Data are collected on the breaking temperature using the *old* process and the *new* process. The data appear below:

	Breaking Temperature		
	New	Old	Difference
	487	475	12
	440	436	4
	495	495	0
	488	483	5
	435	426	9
sample mean	469	463	6
sample SD	28.97	30.27	4.64

For this problem, let μ_{new} be the population mean breaking temperature using the new process and μ_{old} be the population mean breaking temperature using the old process.

(a) [8 points] Suppose the experiment is done using a completely randomized design. That is, a sample of 10 glasses are selected, and randomly assign 5 of them to undergo the new process, and the remaining 5 to undergo the old process.

Test H_0: $\mu_{\text{new}} = \mu_{\text{old}}$ versus H_a: $\mu_{\text{new}} > \mu_{\text{old}}$. Give the test-statistic and P-value and make conclusion using significance level $\alpha = 0.01$. Assume the population SDs are EQUAL ($\sigma_1 = \sigma_2$).

Answer: Since we can assume that the population SDs are equal, we find an estimate for the pooled SD:

$$s_p = \sqrt{\frac{(5-1)839.5 + (5-1)916.5}{5+5-2}} = \sqrt{\frac{7024}{8}} = \sqrt{878} \approx 29.631.$$

The t-statistic is then

$$t = \frac{\overline{x} - \overline{y}}{s_p\sqrt{1/5 + 1/5}} = \frac{469 - 463}{29.631\sqrt{1/5 + 1/5}} \approx 0.320$$

with df $= 5 + 5 - 2 = 8$. Look at the row in the t-table for df $= 8$. The t-statistic 0.320 is less than the smallest value 0.703 in the row, which means the one-side P-value is at least 0.25. So H_0 is NOT rejected.

Based on the results, state what conclusion should be made about whether the new process has a higher average breaking temperature than the old process.

Answer: The data do not provide evidence of a statistically significant difference between the average breaking temperatures of the two methods.

(b) [8 points] Repeat part (a), but assuming the experiment was done using a match-pair design. That is, five plates of glass were selected at random. Each plate was cut in half, with one half undergoing the old process and the other half undergoing the new process.

Answer: The t-statistic is then

$$t = \frac{\overline{d}}{s_d/\sqrt{5}} = \frac{6}{4.64/\sqrt{5}} \approx 2.89$$

with $\mathrm{df} = 5 - 1 = 4$. Look at the row in the t-table for $\mathrm{df} = 4$. The t-statistic 2.891 is between 2.776 and 2.999, which means the P-value is between 0.02 and 0.025. So H_0 is still NOT rejected at $\alpha = 0.01$.

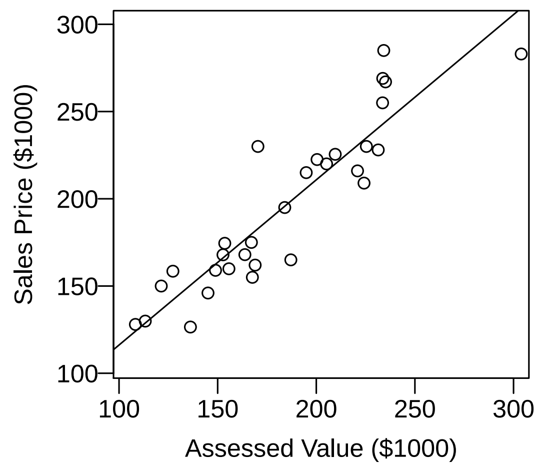
We still fail to reject the hypothesis that the two processes have the same mean breaking temperature, but the P-value is pretty small now. Maybe with a larger sample.

(c) [2 points] In actuality, the data are collected using a matched-pairs design. Explain why the matched-pairs design is more preferable than the completely randomized design. You may (but not necessarily) use the numerical results from parts (a) and (b) to support your answer.

Answer: Paired data have less variability, which can be seen in a much narrower confidence interval in part (b) compared to part (a).

4. [Housing Price] [18 points]

Real estate is typically reassessed annually for property tax purposes. This assessed value, however, is not necessarily the same as the fair market value of the property. An SRS of 30 properties recently sold in a midwestern city was taken. The scatter plot below show the actual sales prices and the assessed values of the 30 properties. Both variables are measured in thousands of dollars.



	Mean	SD
Assessed Value	184.13	45.43
Sales Price	195.84	47.18

Let y_i and x_i be respectively the sales price and the assessed value of the ith property. We use R to fit the regression model

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad \varepsilon_i \sim \text{ i.i.d. } N(0, \sigma).$$

and yield the output below.

```
Call:
lm(formula = Sales.Price ~ Assessed.Value)

Residuals:
    Min      1Q  Median      3Q     Max
-33.649 -12.810   0.206  10.685  47.163

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)    21.49923   15.27936   1.407     0.17
Assessed.Value  0.94682    0.08064  11.741 2.49e-12 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 19.73 on 28 degrees of freedom
Multiple R-squared:  0.8312,    Adjusted R-squared:  0.8251
F-statistic: 137.9 on 1 and 28 DF,  p-value: 2.488e-12
```

(a) [3 points] What are the estimated values for β_0, β_1, and σ?

Answer:

(b) [2 points] Before making further inference, one should check whether the assumptions for the linear regression model appear reasonable for the data. Name one method to check these assumptions.

Answer:

(c) [6 points] Test the hypotheses H_0: $\beta_1 = 1$ versus H_a: $\beta_1 \neq 1$. Together with an insignificant intercept in this model, this would imply that the selling price (y) is equal to the assessed value (x) on average. Give the test statistic, degrees of freedom, and give a range for the P-value. At the 5% significance level, would we reject the null hypothesis?

Answer:

(d) [2 points] For the selling price of three properties currently assessed at respectively \$155,000, \$190,000, and \$285,000, which one can be predicted most accurately? Or they are the same? No need to explain.

Answer:

(e) [5 points] John has a property currently assessed at \$155,000. R predicts its sales value to be \$168,256 along with two different 95% intervals (\$159,447, \$177,065) and (\$126,896, \$209,616). Explain the difference between the two intervals. If John cares about how much his property can sell, which interval better represents the accuracy of the prediction \$168,256?

5. [Football Helmets] [10 points]

A study was conducted at the University of Waterloo on the impact characteristics of football helmets used in competitive high school programs. In the study, a measurement called the Gadd Severity Index (GSI) was obtained on each helmet using a standardized impact test. A helmet was deemed to have failed if the GSI was greater than 1200. Of the 81 helmets tested, 29 failed the GSI 1200 criterion.

(a) [2 point] Assume that the suspension helmets tested were selected at random. What is the estimate of the proportion of suspension helmets that fail the GSI 1200 criterion?

Answer:

(b) [4 points] Based on the sample results, what is the 90% confidence interval estimate for the true population proportion of suspension helmets that would fail the test?

Answer:

(c) [4 points] If the test was to be conducted again, how many suspension-type helmets should be tested so that the margin of error of a 90% confidence interval will not exceed 0.05?

Answer:

6. [STAT220 Review] [10 points]

The following is a list of some statistical methods and techniques discussed in this course.

```
 1. Frequency table         2. Histogram
 3. Stem and leaf plot      4. Boxplot
 5. Scatterplot             6. Correlation
 7. Z confidence interval   8. t confidence interval
 9. Z hypothesis test      10. One sample t test
11. Two sample t test      12. Matched-Pair t test
13. Test for Proportions   14. Chi-squared test
15. Regression
```

For the following situations, select and write the name of the technique that you think is most applicable to the problem described. More than one technique might be acceptable; you only need to list one. If you choose a statistical hypothesis test, also state the null and alternative hypotheses. There are 5 situations; each question is worth 2 points.

(a) A researcher is interested comparing the bone mineral density between female high school athletes and other sedentary high school girls. She has data from a simple random sample of size 80 from each of the two populations, and would like a graphical summary for comparison.

Answer: back-to-back stemplot or side-by-side boxplot

(b) After examining the graph in the previous part, the researcher wants to see whether female high school athletes has lower bone mineral density on average than other sedentary high school girls

Answer: Two-sample t test. Null is means are equal. Alternative is two-sided: athletes has lower mean bone mineral density than other sedentary high school girls

(c) A cereal company conducted a customer survey, to determine if customers of different economic status preferred significantly different size boxes of Chocolate Frosted Sugar Bombs. There were four packaging sizes: small, medium, large, and jumbo. Economic status was divided into lower, middle, and upper classes.

Answer: Chi-squared test. Null is that economic status is independent of packaging size choice. Alternative is that different economic classes have different proportions in size choice.

(d) The cereal company also asked questions about their potential new mascot, a stuffed tiger named Locke. If the survey of roughly 300 people indicates more than 70% favorability, they will begin to use the new mascot.

Answer: Proportion test. Null is $p <= 0.70$ (or $p = 0.70$). Alternative is $p > 0.70$.

(e) A college lecturer believes that attending lecture is very important to success in college. She wants a graph examining grade point average (GPA) against percentage of lectures attended.

Answer: Scatterplot.

7. [Nicotine] [12 points]

A certain brand of cigarettes advertises that the mean nicotine content of their cigarettes is $\mu = 1.5$ milligrams (mg). To test this, a random sample of 100 cigarettes of this brand were examined and the P-value for testing $H_0 : \mu = 1.5$ mg versus H_a: $\mu \neq 1.5$ mg was found to be = 3.2%.

(a) [2 points] True or False and explain briefly: The value 1.5 mg is in the 99% confidence interval for the actual mean nicotine content of cigarettes of this brand.

Answer:

(b) [2 points] True or False and explain briefly: A P-value of 3.2% means the probability that H_0 is true is 3.2%.

Answer:

(c) [2 points] True or False and explain briefly: The P-value for the one sided alternative H_a: $\mu < 1.5$ mg is $3.2\%/2 = 1.6\%$.

Answer:

(d) [2 points] Explain what is a Type II error in the test $H_0: \mu = 1.5$ mg versus H_a: $\mu \neq 1.5$ mg.

Answer:

(e) [4 points] A public health organization concerns about the excessive amount of nicotine in the cigarettes and wish to do a one-sided test $H_0: \mu = 1.5$ mg versus H_a: $\mu > 1.5$ mg. It takes a random sample of 100 cigarettes of this brand for examination, and decides to reject the null if the mean nicotine content of the 100 cigarettes exceeds 1.55 mg.

Suppose the nicotine content in cigarettes of this brand is Normally distributed with standard deviation $\sigma = 0.2$ mg. What is the power of this test if the actual value of μ is 1.55 mg?

Answer: