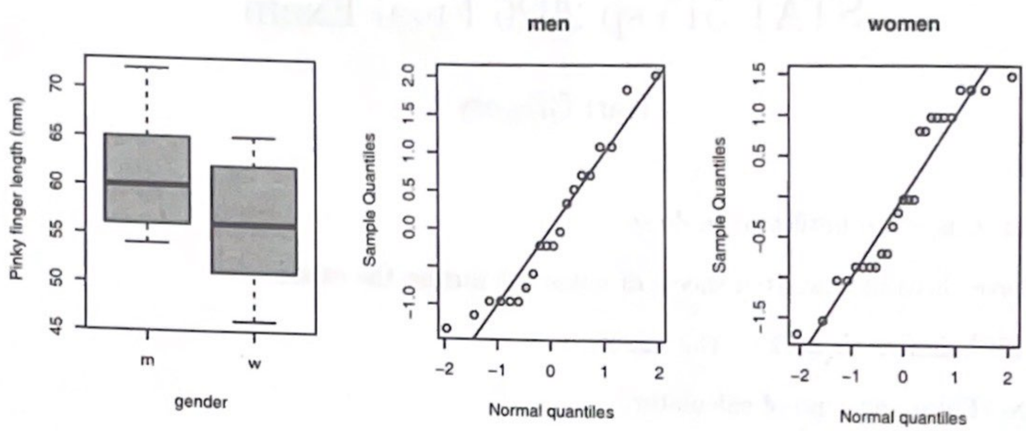
STAT 515 sp 2026 Final Exam

Karl Gregory

- Do not open this exam until told to do so.
- You may have three handwritten sheets of notes out during the exam.
- You have 150 minutes to work on this exam.
- You may NOT use any kind of calculator.

A z-table is appended to this exam.

1. The pinky finger lengths and genders of several statistics students were recorded. Below are some plots of the data and some R output.




```
Two Sample t-test

data:  pnk by gender
t = 2.885, df = 43, p-value = 0.006096
alternative hypothesis: true difference in means
between group m and group w is not equal to 0
95 percent confidence interval:
 1.489776 8.410224
sample estimates:
mean in group m mean in group w
          61.15           56.20
```

(a) Explain the purpose of looking at the two plots on the right.

The purpose of the two Normal quantile-quantile plots on the right is to ensure that the samples being tested came from a Normal population so that the samples can be treated as having a Normal distribution. The two plots make sure that the data from the men and the data from the women can be treated as being Normally distributed, which does appear to be the case as the points on both plots appear to follow a linear trend. Such also informs the specific test statistic to calculate and value to which it should be compared for hypothesis testing.

(b) Write down the null and alternate hypotheses one should test in order to see if the average pinky finger length of men exceeds that of women. μ_m - average pinky length of men, μ_f - average pinky length of women

$H_0: \mu_m - \mu_f \leq 0$

$H_1: \mu_m - \mu_f > 0$

(c) Based on the R output, give the value of the test statistic for testing the hypotheses from part (b).

$T_{test} = 2.885$

(d) Give the p value for testing the hypotheses from part (b).

$$\text{p-value} = P(T > T_{test}) = \frac{0.006096}{2} = \boxed{0.003048}$$

(e) Based on the R output, what is your conclusion about the mean pinky finger lengths for men and women?

As the p-value is very small, we can reject H_0 at $\alpha = 0.05$ and even $\alpha = 0.01$ to conclude that the average pinky length of men does likely exceed that of women.

2. Suppose the orders of one hundred diners are recorded. Sixty of them order a salad, and, among these sixty, twenty order a dessert. Among the forty who do not order a salad, the number who order a dessert is also twenty.

(a) Summarize the counts in a two-by-two table. Include row and column sums in the margins.

Salad

Dessert	yes	no	TOTAL
yes	20	20	40
no	40	20	60
TOTAL	60	40	100

(b) It is of interest to test whether there is an association between ordering a salad and ordering a dessert. Write down the corresponding null and alternate hypotheses.

H_0: There is no association between ordering a salad and dessert.

H_1: There is an association between ordering a salad and dessert.

(c) Give the table of expected counts under the null hypothesis of no association.

Salad

Dessert	yes	no
yes	24	16
no	36	24

(d) Is the rule of thumb satisfied for using the chi-squared test of association? State the rule and explain why it is or is not satisfied.

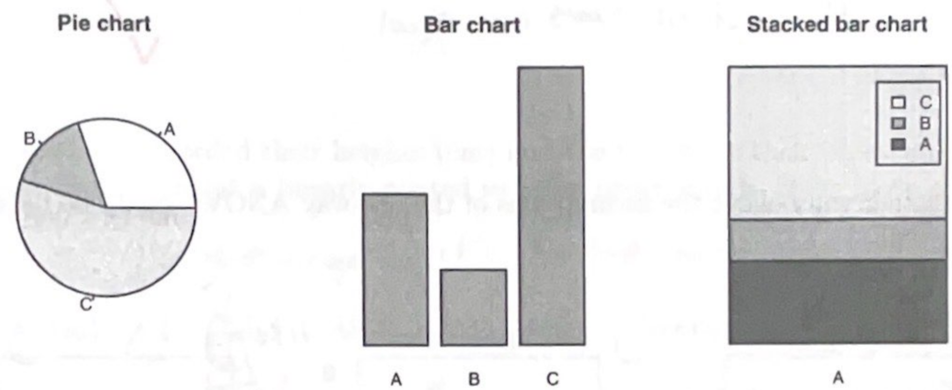
The rule of thumb for using the χ^2 test of association is that all expected values E_{ij} for all i,j must be greater than or equal to 5. This rule is satisfied

(e) Pearson's test statistic for comparing the observed to the expected counts has the value 2.78. Carefully explain how we come to a decision whether to reject or not to reject the null hypothesis based on the value of the test statistic.

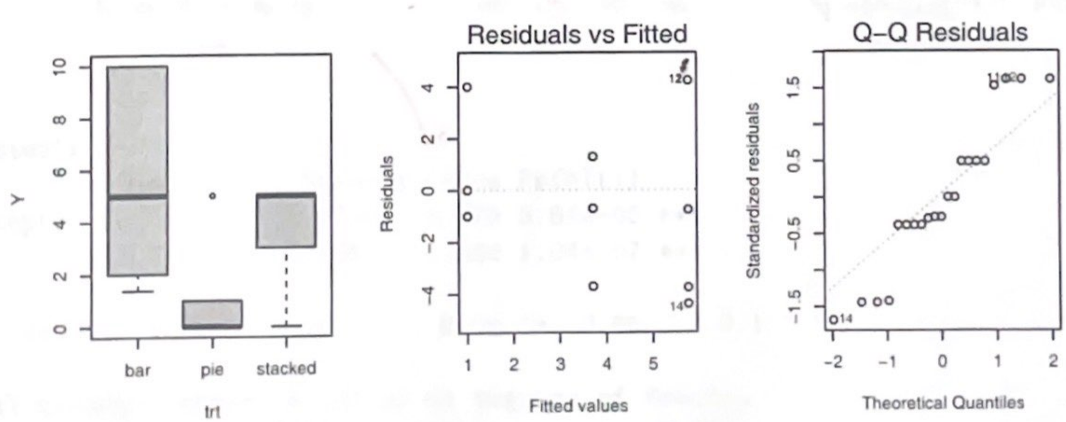
Reject H_0 if $W_{test} = 2.78 > \chi^2_{(R-1)(C-1),\alpha}$

where $R=2$ and $C=2$, and α is the significance level we decide.

3. Each student in a statistics class was shown one of the three charts below and allowed 5 seconds to guess the percentage of the whole made up by category B:



The absolute values of the differences between the students' guesses and the true percentage made up by category B (which was 15%) were recorded as response values in a one-way ANOVA model. Some plots and output are shown below:



Analysis of Variance Table

Response: Y

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
trt	2 ?	80.05	40.025	5.0211	0.0185 *
Residuals	18	143.49	7.971		

$MS_T = \frac{SS_{treat}}{df_{treat}}$

$df_{trt} = \frac{SS_{trt}}{MS_{trt}} = \frac{80.05}{40.025} = 2$

$\frac{40.025}{7.971}$

(a) Replace the question mark in the ANOVA table with the correct degrees of freedom value. ? = 2

(b) Give the null and alternate hypotheses for which the F value in the ANOVA table is the test statistic.

$H_0: \mu_A = \mu_B = \mu_C$

H_1: not all means are equal

(c) State whether you believe the assumptions of the one-way ANOVA model to be satisfied. Explain your answer.

I dont believe the assumptions are satisfied. When looking at the box plot, we see the boxes are not close to the same size at all, along with the residuals - Fitted plot show large deviations around the mean that arent all the same. Those two things tell me unequal variation, so that assumption isnt met.

When looking at the Q-Q plot, there are clear outliers in the data. and the data does not appear to follow normal distributi That assumption isnt met either

(d) How does one obtain the value 5.0211 from the other values in the ANOVA table?

That is the F-statistic which is

$$= \frac{\text{Mean Square Treat}}{\text{Mean Square Error}} \rightarrow \frac{40.025}{7.971} = 5.0211$$

(e) State your conclusions about the hypotheses in part (b) at the $\alpha = 0.05$ significance level.

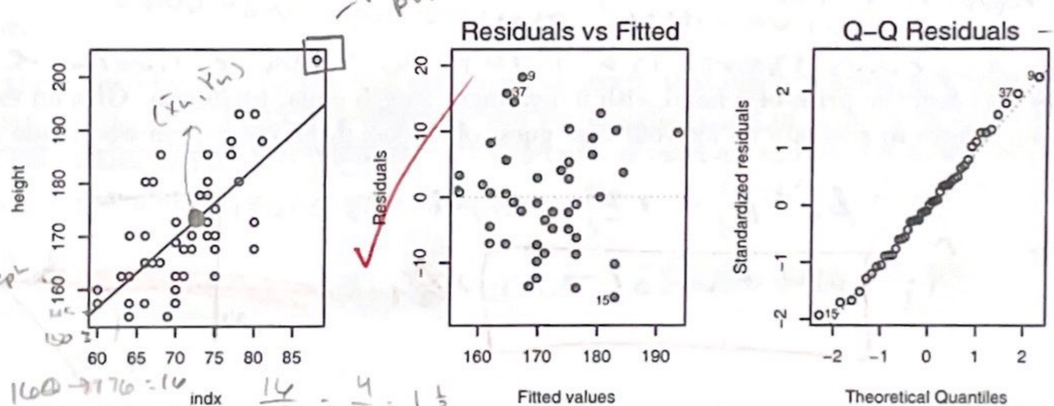
p-value = 0.0185

at a 0.05 significance level, we reject the null hypothesis with a p-value of 0.0185

4. Several statistics students recorded their heights (cm) and the lengths of their index fingers (mm). It is of interest to see whether height is linearly related to index finger length. Some plots and output are below.

- leverage point

- normal



y-intercept ≈ 154

160 → 176 = 16

63 → 75 = 12

$\frac{16}{12} = \frac{4}{3} = 1\frac{1}{3}$

1.33

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  78.0144    14.7773   5.279 3.81e-06 ***
indx          1.3116     0.2067   6.346 1.04e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8.228 on 44 degrees of freedom
Multiple R-squared:  0.4779,Adjusted R-squared:  0.466
```

small p-val

positive relationship

```
F-statistic: 40.27 on 1 and 44 DF,  p-value: 1.045e-07

                 2.5 %      97.5 %
(Intercept) 48.2327000 107.796061
indx         0.8950504   1.728079
```

(a) What proportion of variability in height can be explained by considering index finger length?

Multiple R squared: 0.4779

(b) State whether you believe the assumptions of the simple linear regression model are satisfied. Explain your answer.

Yes, residuals are random and the Q-Q Plot is approximatly linear

(c) Suppose you find the print of a hand with index finger length equal to 76 mm. Give an expression (you do not have to evaluate it) for your best guess of the height of the person who made the hand print.

78.0144 + 1.3116(76)

(d) What if you find the print of a hand with index finger length equal to 50 mm?

It is extrapolation So it is not reliable

(e) State the null and alternate hypotheses for which the number 1.04×10^{-7} (which appears in the output) is the p value.

$H_0: \beta_1 = 0$

$H_1: \beta_1 \neq 0$

(f) Give an interpretation of the confidence interval [0.895, 1.728] which appears in the output (I have rounded the numbers).

we are 95% confident the

the slope is between 0.895 and 1.728

(g) Draw a dark circle on the scatterplot at your best guess of the location $(\bar{x}_n, \bar{Y}_n)$, where $\bar{x}_n$ is the mean of the observed index finger lengths and $\bar{Y}_n$ is the mean of the observed heights.

(h) Draw a small box around the point in the scatterplot which you believe has the highest leverage value.

5. Suppose there are 5 bowling balls which are identical except that one is magical and delivers, no matter what, a strike with probability 3/4. Suppose you get a strike 1 out of 4 times on average when using non-magical bowling balls. You select one of the 5 balls at random and send it down the lane...

(a) Give the probability that you get a strike.

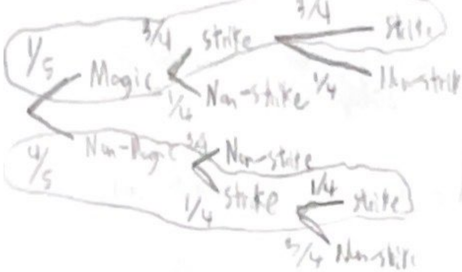
$P(\text{strike}) = P(m)P(S|m) + P(NM)P(S|NM)$

$= (\frac{1}{5})(\frac{3}{4}) + (\frac{4}{5})(\frac{1}{4}) = \boxed{\frac{7}{20}}$

(b) Given that you got a strike, what is the probability you chose the magic bowling ball?

$P(m|s) = \dfrac{P(S|m)P(m)}{P(s)} = \dfrac{\frac{3}{20}}{\frac{7}{20}} = \boxed{\frac{3}{7}}$

(c) Suppose you choose a ball and with the same ball you get two strikes in a row. What is the probability that you chose the magic ball?



$P(2\text{ Strikes}) = \frac{3}{20} \cdot \frac{3}{4} + \frac{4}{20} \cdot \frac{1}{4}$

$\frac{9}{80} + \frac{4}{80} = \frac{13}{80}$

$P(\text{Magic} \mid 2\text{ Strikes}) = \frac{P(\text{Magic} \cap 2\text{ strikes})}{P(2\text{ strikes})}$

$= \frac{\frac{9}{80}}{\frac{13}{80}} = \frac{9}{13}$

6. Suppose the index finger lengths of women are normally distributed with mean 56 mm and standard deviation 5 mm.

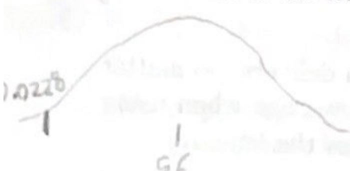
(a) Give the probability that the index finger length of a randomly selected woman exceeds 63.5 mm.

$z = \frac{63.5 - 56}{5} = \frac{7.5}{5} = \frac{3}{2} = 1.5$

using z-table $\;\; 0.50010 - 0.4332 = 0.0668$

0.0668

(b) A woman claims that the length of her index finger is at the 2.28 percentile of index finger lengths. What is the length of her index finger?



$0.50010 - 0.0228 = 0.4772$

z-table, $z = -2$

$X = \mu + \sigma z$

$X = 56 + 5 \cdot -2$

$X = 56 - 10 = 46$

46 mm

(c) Give the probability that the mean of the index finger lengths of sixteen randomly selected women will be less than 54.75 mm.

$Z = \frac{\bar{X}_n - \mu}{\sigma / \sqrt{n}} = \frac{54.75 - 56}{5/4} = \frac{-1.25}{\frac{5}{4}} = \frac{-\frac{5}{4}}{\frac{5}{4}} = -1$

$0.50010 - 0.3413 = 0.1587$

0.1587