STAT 515 fa 2021 Final Exam

Karl B. Gregory

- Do not open this exam until told to do so.
- You may have 3 handwritten sheet of notes out during the exam.
- You have 2.5 hours to work on this exam.
- You may use a simple calculator.
- If you are unsure of what a question is asking for, do not hesitate to ask me for clarification.
- *Good luck, and may the odds be ever in your favor!*

Some upper quantiles of some chi-squared distributions:

α	0.10	0.05	0.025	0.01	0.005
$\chi^2_{1,\alpha}$	2.71	3.84	5.02	6.63	7.88
$\chi^2_{2,\alpha}$	4.61	5.99	7.38	9.21	10.60
$\chi^2_{3,\alpha}$	6.25	7.81	9.35	11.34	12.84
$\chi^2_{4,\alpha}$	7.78	9.49	11.14	13.28	14.86

$$S^2_{\text{pooled}} = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}, \quad T_{\text{test}} = \frac{\bar{X}_1 - \bar{X}_2}{S_{\text{pooled}}\sqrt{1/n_1+1/n_2}}, \quad \bar{X}_1 - \bar{X}_2 \pm t_{n_1+n_2-2} S_{\text{pooled}}\sqrt{1/n_1+1/n_2}$$

For $X \sim \text{Binomial}(n,p)$ we have $P(X=x) = \binom{n}{x} p^x (1-p)^{n-x}$.

1. In the USA, approximately 1 out of 700 babies is born with Trisomy 21 (Down syndrome). About 1 in 4 babies born with Trisomy 21 have a heart defect called atrioventricular septal defect. Across the entire population, roughly 1 in 1,860 babies are born with this heart defect. These figures are taken from the following sources:

   - https://www.cdc.gov/ncbddd/heartdefects/avsd.html
   - https://www.cdc.gov/ncbddd/birthdefects/downsyndrome.html
   - https://www.mottchildren.org/conditions-treatments/ped-heart/conditions/atrioventricular-septal-defect

   Consider a randomly selected newborn:

   (a) Given the presence of the atrioventricular septal defect, give the probability of Trisomy 21.

   (b) Given the absence of Trisomy 21, give the probability of the atrioventricular septal defect.

2. A survey of the marital and employment statuses of a sample of males aged 35–44 resulted in the table:

	Married	Widowed, divorced, or separated	Never married
Employed	638	133	102
Unemployed	27	8	6
Not in labor force	35	12	20

It is of interest whether there is an association between marital and employment statuses (See pg. 490 of [2]).

(a) Give the bottom row of the table of expected values under the null hypothesis of no association.

(b) Which entry in the bottom row of the observed table will make the greatest contribution to the value of the test statistic $W_{\text{test}} = \sum_{j=1}^{J} \sum_{k=1}^{K} (O_{jk} - E_{jk})^2 / E_{jk}$?

(c) Give the $\alpha = 0.05$ critical value for rejecting the null hypothesis (use table on front of exam).

(d) Determine whether or not to reject the null hypothesis at the $\alpha = 0.05$ significance level based on the test statistic $W_{\text{test}} = \sum_{j=1}^{J} \sum_{k=1}^{K} (O_{jk} - E_{jk})^2 / E_{jk}$.

3. An experiment conducted in Nigeria compared the chlorophyll content (an indexed reading from -9.9 to 199.9 from a device clamped onto a leaf for two seconds) in the leaves of 145 maize seedlings, of which by random assignment 36, 27, 39, and 43 were treated with 0, 5, 10, and 20 grams, respectively, of an N-P-K (nitrogen-phosphorus-potassium) fertilizer [1]. The experiment resulted in the values $\mathrm{SS}_{\mathrm{Total}} = 21{,}815.2$ and $\mathrm{SS}_{\mathrm{Treatment}} = 11{,}373.06$.

   (a) Write down the hypotheses of interest to the researchers.

   (b) Give the values (i)–(vi) that are missing from the ANOVA table:

Source	Sum of Sq	df	Mean Sq	F	p-value
Treatment	11373.06	(i)	(ii)	(vi)	
Error	(iii)	(iv)	(v)		
Total	21815.2	144			

   (c) The table below gives some upper quantiles of the F-disribution relevant to this experiment:

α	0.10	0.05	0.025	0.01	0.005
$F_{\text{(i)},\text{(iv)},\alpha}$	2.12	2.67	3.21	3.92	4.46

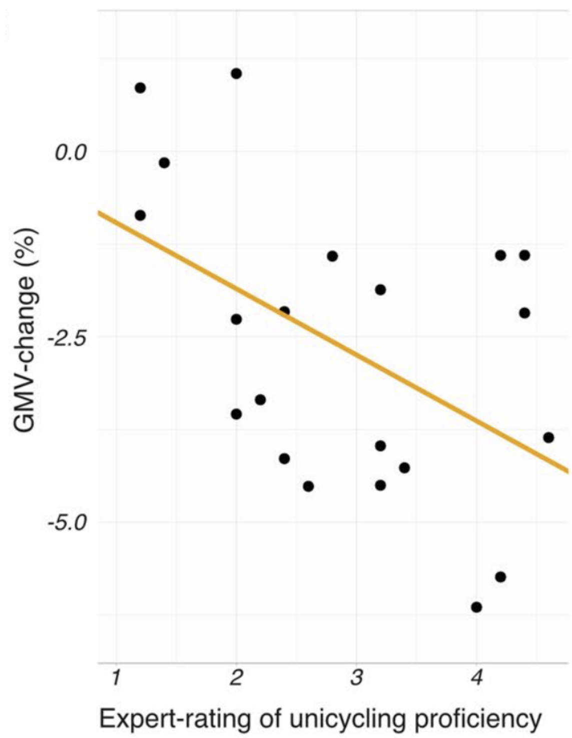
   State whether you think the null hypothesis should be rejected and explain why.

   (d) If you had access to the entire data set:

      i. How could you check the assumption of equal variances among the treatment groups?

      ii. How could you check the assumption that the responses are Normally distributed around the treatment means?

4. In a study of neuroplasticity—the ability of the brain to reorganize itself in response to environmental demands—several female subjects with no prior unicycling experience were trained by semi-professional unicyclists for three weeks, before and after which their brains were imaged [3]. The figure below plots the percent changes in gray matter volume at the right superior temporal gyrus (a region of the brain) from before to after the training against expert ratings of post-training unicycling proficiency for the subjects in the study.



The authors reported a p-value equal to 0.030 for the significance of the relationship between percent change in gray matter volume and the expert rating of unicycling proficiency.

(a) Write down a simple linear regression model for the relationship between the percent change in gray matter volume and the expert rating of post-training unicycling proficiency. Define your variables and your parameters.

(b) Do your best to ascertain (approximately) the values of the least-squares estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ of the regression coefficients by studying the plot.

(c) What difference in the percent change of gray matter volume is associated with an increase of 1 point in a subject's unicycling proficiency rating?

(d) Write down the null and alternate hypotheses to which the reported p-value of 0.030 corresponds.

(e) Identify two points on the scatter plot (label them) which have a strong influence on the least-squares regression line. Explain why you selected these two points (there is more than one correct choice of two points).

(f) The authors reported one of the following values for the correlation between the percent change in gray matter volume and the expert rating of post-training unicycling proficiency. Which one is it?

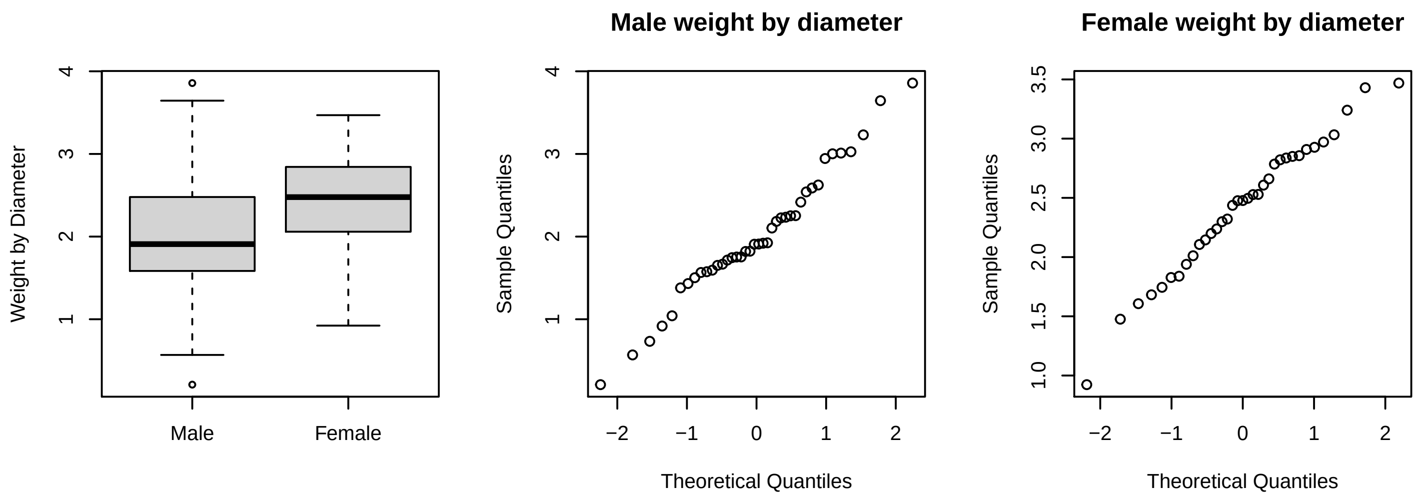
A. $r_{xY} = -0.030$
B. $r_{xY} = -1.101$
C. $r_{xY} = -0.876$
D. $r_{xY} = -0.473$
E. $r_{xY} = 0.473$
F. $r_{xY} = 1.101$
G. $r_{xY} = 0.030$

(g) What percent change of gray matter volume would you expect for subjects with an expert rating of unicycling proficiency equal to 6?

5. From a large data set with measurements on 1,527 male abalones and 1,307 female abalones, 40 males and 35 females were sampled. The whole weight divided by the diameter of each of these was recorded. The means and standard deviations of the weight-by-diameter values for the two samples were

$$\bar{X}_M = 2.006, \quad \bar{X}_F = 2.420, \quad S_M = 0.787, \quad \text{and} \quad S_F = 0.572,$$

where M and F denote male and female, respectively. Below are boxplots and Normal quantile-quantile plots of the samples.



Suppose it if of interest whether there is a difference in the weight-by-diameter ratio of male and female abalones.

(a) Write down the hypotheses of interest (male minus female).

(b) Give the value of the test statistic for testing the hypotheses of interest; assume equal variances.

(c) Use the t-table to determine a range within which the p-value lies.

(d) Give a 95% confidence interval for the true difference in weight-by-diameter ratio means. If you cannot find the row of the t-table you need, just use the row above it.

(e) State whether you should conclude, based on these data, that the weight-by-diameter means differ.

6. You are hosting 8 friends for an afternoon gathering and you are wondering how many cups of coffee to brew. For each guest, the probability that he or she will want coffee is $3/5$. Find the probability of the events (evaluate your expressions):

   (a) All 8 guests want coffee.

   (b) At least one guest wants coffee.

   (c) Exactly 6 guests want coffee.

   (d) No more than 6 guests want coffee.

References

[1] Hussaini Abubakar, Haruna Danyaya Abubakar, and Aminu Salisu. One way anova: Concepts and application in agricultural system.

[2] David Freedman, Robert L. Pisani, Roger Purves, and Ani Adhikari. *Statistics, Second Edition*. W.W. Norton & Company, Inc., 1991.

[3] Ilona Papousek, Bernhard Weber, Karl Koschutnig, Andreas Schwerdtfeger, Christian Rominger, Elisabeth M Weiss, Markus Tilp, and Andreas Fink. Learning unicycling evokes manifold changes in gray and white matter networks related to motor and cognitive functions. *Scientific Reports*, 9(1):4324–4324.