Stat 500 Solutions to Homework #6 Fall 2020

Common mistakes: please see inside the solutions(the red parts)

1. 10 points

(a) i. The minimum distance between two sample means required to reject the hypothesis that two trials have the same population means, using an experimentwise Type I error level of $\alpha = .05$ is

$$HSD = \frac{1}{\sqrt{2}} q(5, 95, 0.95)\sqrt{MSE(\frac{2}{n})} = 65.28$$

ii. The following display indicates which means are significantly different at the experimentwise $\alpha = .05$ level by underlining sets of means that are not significantly different.

First Trial	Second Trial	Third Trial	Fifth Trial	Fourth Trial
909	856	845	831.5	820.5

So $\mu_1 - \mu_4$ and $\mu_1 - \mu_5$ are significantly different.

(b) There are a total of 10 comparisons, and the type I error rate that we need to control for each comparison so that the experiment-wise type I error rate is 0.05 is $0.05/10 = 0.005$.

The type I error rate for Bonferonni is $1 - \alpha/m$, not $1 - \alpha/2m$

(c) SAS code for doing this is:

```
proc glm data=set1; class trial;
model velocity = trial / p;
contrast "1 vs 2345" trial -4 1 1 1 1;
estimate "1 vs 2345" trial -4 1 1 1 1 / divisor=-4;
contrast "2 vs 345" trial 0 -3 1 1 1;
estimate "2 vs 345" trial 0 -3 1 1 1 / divisor=-3;
contrast "3 vs 45" trial 0 0 -2 1 1;
estimate "3 vs 45" trial 0 0 -2 1 1 / divisor=-2;
contrast "4 vs 5" trial 0 0 0 1 -1;
estimate "4 vs 5" trial 0 0 0 1 -1;
means trial /alpha=.05 bon lsd scheffe tukey snk hovtest=bf;
output out=set1 residual=resid predicted=ymean;
run;
```

The results of the estimate statements are:

Contrast	Estimate	Standard Error	t value	Pr > |t|
1 vs 2345	70.7500000	18.5584071	-3.81	0.0002
2 vs 345	23.6666667	19.1670404	-1.23	0.2200
3 vs 45	19.0000000	20.3297164	-0.93	0.3524
4 vs 5	-11.0000000	23.4747345	0.47	0.6404

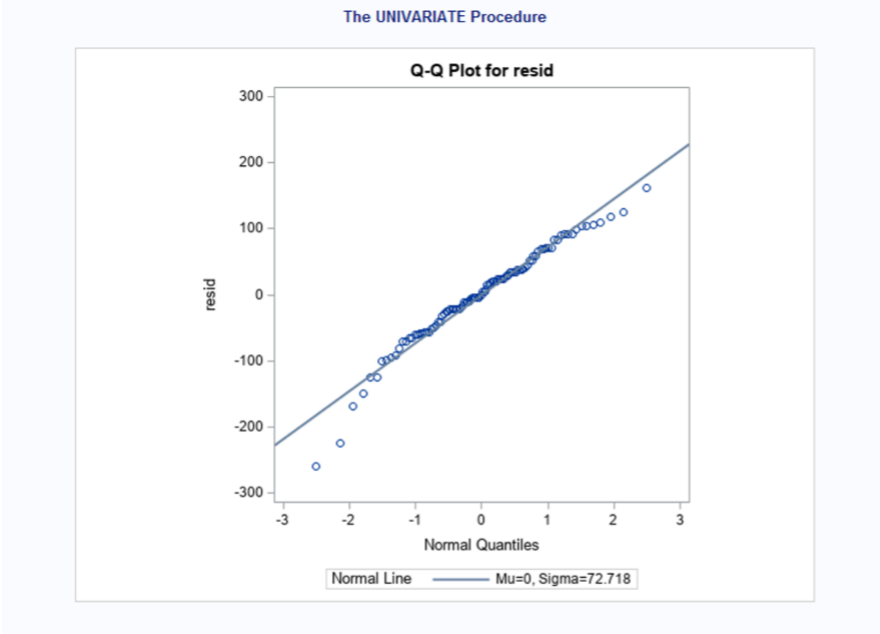
Based on the t tests and associated p-values, we have strong evidence to conclude that the mean speed measured on the first trial is larger than the average of the mean speeds for the other four trials. This analysis does not reveal any significant differences between means for the last four trials.

(d)

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
1 vs 2345	1	80089.0	80089.0	14.53	0.0002
2 vs 345	1	8401.7	8401.7	1.52	0.22
3 vs 45	1	4813.3	4813.3	0.87	0.3524
4 vs 5	1	1210.0	1210.0	0.22	0.6404

$80089 + 8401.67 + 4813.33 + 1210 = 94514$ is equal to the Sum of Squares between trials, which is $20 * [(909 - 852.4)^2 + (856 - 852.4)^2 + (845 - 852.2)^2 + (831.5 - 852.4)^2 + (820.5 - 852.4)^2]$.

(e)



The Q-Q plot of residuals is shown above. This plot suggests that the random errors could be sampled from a left skewed population. The value of the Shapiro-Wilk test is 0.9678 with p-value $0.0150 < 0.05$, which is sufficient evidence to reject the null hypothesis that the random errors were sampled from a normal distribution. This is consistent with the pattern observed in the Q-Q plot. Therefore, the assumption of normality for the measurements of the speed of light in air does not appear to be valid, although this the skewness in the error distribution may not be enough to severely distort the p-value

obtained by comparing the F-statistic for testing the null hypothesis of equal means to a central F-distribution (consider your results in part d).

(f)

$$\begin{cases} H_0: & \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 = \sigma_5^2 \\ H_1; & Not\ all\ variances\ are\ equal \end{cases}$$

From the Brown-Forsythe test, we obtain that p-value = 0.1622 > 0.05; thus, we fail to reject the null hypothesis at the significance level of 0.05, i.e., the data do not provide sufficient evidence to support the claim that the population variances are heterogeneous.

(g)

Kruskal-Wallis Test Chi-Square	15.0221
DF	4
Pr > Chi-Square	0.0047

Based on the Kruskal-Wallis test and the associated p-value 0.0047, we can reject the null hypothesis that the light speed measurements for all five trials were sampled from the same distribution. The sum of the ranks for the first trial were much larger than the expected value under the null hypothesis, indicating that the first trial tended to produce larger values for the speed of light measurements than some of the other trials.

2. 10 points

(a)

$$\begin{cases} H_0: & \frac{1}{6}\sum_{i=1}^{6} E(log(entrance\ area\ for\ species\ i)) = \frac{1}{3}\sum_{j=7}^{9} E(log(entrance\ area\ for\ species\ j)) \\ H_a: & \frac{1}{6}\sum_{i=1}^{6} E(log(entrance\ area\ for\ species\ i)) \neq \frac{1}{3}\sum_{j=7}^{9} E(log(entrance\ area\ for\ species\ j)) \end{cases}$$

The total sample size is N = 294 and the number of groups is r = 9.

Under H_0

$$\frac{|\sum_{i=1}^{9} c_i n_i^{-1} \sum_{k=1}^{n_i} log(Y_{ik})|^2}{(r-1)MS_{error}\sum_{i=1}^{9}\frac{c_i^2}{n_i}} \sim F_{(r-1,N-r)}$$

The test statistic is $\frac{|(\frac{7.347+7.369+7.428+7.487+7.563+7.568}{6} - \frac{8.214+8.272+8.297}{3})|^2}{(9-1)\times 0.1919\times(\frac{1}{36}\times(\frac{1}{127}+\frac{1}{44}+\frac{1}{24}+\frac{1}{41}+\frac{1}{18}+\frac{1}{16})+\frac{1}{9}\times(\frac{1}{11}+\frac{1}{7}+\frac{1}{6}))} = 8.2573$

The critical value is $F_{8,285,0.95} = 1.97$

The p-value for the observed data is

$$p - value = P(F_{(8,285)} > 8.2573) = 4.3758e^{-10}$$

Since the p-value is less than 0.05, we reject null hypothesis significantly at the significance level $\alpha = 0.05$, and we conclude that the means of log(entrance area) are significantly different for the two sets of species.

- You all knew which formula to use, i.e. the one in notes page 354. But that one is comparing pairs of group means, here you are dealing a contrast so adjustment is necessary. You need to come up with the variance formula for the contrast.

- Cannot simply add up all sample sizes in each group, i.e. use $\sum_{i=1}^{6} n_i = 270$ and $\sum_{i=7}^{9} n_i = 24$ as the n_i, n_j in the formula.

(b) $C.I_{Sche} = \sum_{i=1}^{9} c_i n_i^{-1} \sum_{k=1}^{n_i} log(Y_{ik}) \pm \sqrt{(r-1)F_{r-1,N-r,1-\alpha}} \sqrt{MS_{error} \sum_{i=1}^{9} \frac{c_i^2}{n_i}} = (-1.1914, -0.4099)$
Do NOT use $1 - \frac{\alpha}{2}$ in F distributions. Unlike the "two tail" stuff in the two sided t test, F distribution is non-negative and non-symmetric.

(c) $C.I_t = \sum_{i=1}^{9} c_i n_i^{-1} \sum_{k=1}^{n_i} log(Y_{ik}) \pm t_{N-r,1-\alpha/2} \sqrt{MS_{error} \sum_{i=1}^{9} \frac{c_i^2}{n_i}} =$ (-.9944, -.6070)
Confidence interval based on an ordinary t-test is 50% narrower than the confidence interval based on the Scheffe' method. This is to be expected, because the Scheffe' method provides simultaneous 95% confidence for intervals made for an infinite number of contrasts.

(d) Although only one confidence interval was evaluated and reported, the researchers first explored the data to determine which confidence interval to make. They mentally explored an unspecified number of contrasts before selecting the one that they finally evaluated. In essence, their "data snooping" considered an unspecified number of potential confidence intervals and the Scheffe' method is needed.

3. 6 points

(a) A randomized block design (RCBD). Experimental units are plots, and blocks are fields.

Source of variation	df
Fields	4
Densities	5
Error	20
Corrected total	29

Table 1: 4.a

(b) A completely randomized design (CRD) with circuit boards as experimental units and no blocks. The three treatments are the three cutting speeds.

Source of variation	df
Treatments	2
Error	12
Corrected total	14

Table 2: 4.b

(c) A repeated measures experiment. Repeated measures are taken on the productivity of each plant with five different types of music. The experimental units are the 8 plants. Responses taken at the same plant under different music conditions may be positively correlated, i.e. highly productive plants will perform better than less productive plants for each type of music. If you are convinced that any two responses taken at a single plant have the same positive correlation, you can analyze these data using the ANOVA

Plants (blocks)	7
Types of music	4
Error	28
Corrected total	39

Table 3: 4.c

for a randomized block design where you treat the plants as blocks and the observations under the five different types of music as five runs performed in the same block.

4. 11 points

(a) Blocks are patients, and the treatments in this experiment are the placebo and the drug given.

(b) $H_0 : \mu_D = 0$ versus $H_a : \mu_D > 0$

Using a paired t-test, the test statistic is $t = 1.326$. df=9, and p-value= Pr $(t_9 > 1.326) = 0.1088$.

Recall what you learned in paired t test. The denominator should be $sd(\bar{X}) = \frac{sd(X)}{n}$ instead of $sd(X)$.

(c) No, because the p-value is larger than 0.05.

(d) W = 0.9258; p-value = 0.4079

Based on the test statistic of 0.9258 and a corresponding p-value of 0.4079, there is not enough evidence to reject the null hypothesis. We can conclude that the differences in observed drug and placebo responses are consistent with those from a normal distribution.

(e) (The observation with difference value of 0 was eliminated before calculating the ranks and n)

$Z = \frac{|31-9\times10/4|-0.5}{\sqrt{9\times10\times19/24}} = 0.9477582$ One-sided p-value= $Pr(z > Z) = 0.1716$

Since the p-value is greater than 0.05 we do not reject the null hypothesis and conclude that there is no evidence that the mean of the drug group is greater than the mean of the placebo group.

I don't know whether you've covered this adjustment for ties in class so I gave credit to both 9 and 10. Just keep in mind that ties should be ignored when doing Wilcoxon signed rank test.

(f) $H_0 : \Pr(+) = \Pr(-) = 0.5, H_A = \Pr(+) \neq 0.5$. Table 4 shows that there are $S = 4$ pairs of negative differences out of 9 pairs (ties are ignored).

p-value is

$$\left[\binom{9}{0}0.5^9 + \binom{9}{1}0.5^9 + \binom{9}{2}0.5^9 + \binom{9}{3}0.5^9 + \binom{9}{4}0.5^9\right] = 0.5$$

There is no difference on the effects of hours of sleep per night between the drug and placebo.

Pair	Difference	sign
1	1.3 - 0.6 = 0.7	+
2	1.1 - 1.1 = 0	
3	6.2 - 2.5 = 3.7	+
4	3.6 - 2.8 = 0.8	+
5	4.9 - 2.9 = 2	+
6	1.4 - 3.0 = -1.6	-
7	6.6-3.2 = 3.4	+
8	4.5 - 4.7 = -0.2	-
9	4.3 - 5.5 = -1.2	-
10	6.1 - 6.2 = -0.1	-

Table 4: Sign test.

If you are doing 2-sided test you will get p-value of 1, which is reasonable because all possible situations are "same or more extreme" than $\begin{pmatrix} 9 \\ 5 \end{pmatrix}$

(g) A 95% confidence interval is computed as $(\bar{Y}_{drug} - \bar{Y}_{placebo}) \pm t_{n-1,0.975}\sqrt{\frac{S^2_{diff}}{n}}$.

To have a confidence interval with width equal to 0.75 hours requires $0.75 = 2t_{n-1,0.975}\sqrt{\frac{S^2_d}{n}}$ which implies that $\sqrt{n} = \frac{2t_{n-1,0.97}S_d}{0.75} \Rightarrow n = \frac{4t^2_{n-1,0.97}S^2_d}{(0.75)^2} \approx \frac{4t^2_{n-1,0.975}(3.2006)}{(0.75)^2}$.

Using the upper 0.975 percentile of the standard normal distribution, an approximation to the sample size is

$n_0 = \frac{4Z^2_{0.975}S^2_d}{(0.75)^2} \approx \frac{4(1.96)^2(3.2006)}{(0.75)^2} = 87.43412 \Rightarrow 88$ patients

Update using a percentile from a T-distribution to obtain

$n = \frac{4t^2_{n_0-1,0.975}S^2_d}{0.75^2} \approx \frac{4(2.079614)^2(3.2006)}{0.75^2} = 89.91465 \Rightarrow 90$ patients

Use the formula in notes page 159-162. Again you will have to do adjustments (in the standard deviation part), so you must understand how this formula is derived.

5. 3 points

Using the binomial distribution Y~Bin(25,0.5), implied by the null hypothesis, the exact p-value is $2Pr(Y >= 20) = 2\left[\begin{pmatrix} 25 \\ 20 \end{pmatrix}0.5^{25} + \begin{pmatrix} 25 \\ 21 \end{pmatrix}0.5^{25} + ... + \begin{pmatrix} 25 \\ 24 \end{pmatrix}0.5^{25} + \begin{pmatrix} 25 \\ 25 \end{pmatrix}0.5^{25}\right] =$ 0.0041. The normal approximation to the binomial distribution yields $Z = \frac{20-25\times0.5-0.5}{\sqrt{25\times0.5\times0.5}} = 2.8$ and an approximate p-value of 0.0052. Both methods lead to the same conclusion that a higher proportion of patients experience fewer side effects when THC is used instead of compazine.

Ties MUST be eliminated

6. 10 points

(a) ANOVA

Source	df	SS	MS	F	Pr>F
Blocks	4	1.419	0.355	146.89	<0.0001
Diets	2	1.320	0.660	273.55	<0.0001
Error	8	0.019	0.002		
Corrected total	14	2.759			

(b) Based on the F-statistic of 273.55 and an associated p-value of <0.0001, we reject the null hypothesis of equal mean plasma lipid levels for all three diets and conclude that the mean plasma lipid level for at least one of the diets differs from the others.

(c) Using Tukey's HSD procedure, we can conclude that each diet has a significantly different mean plasma lipid level. As the fat content of the diet increases, the mean plasma lipid level decreases.

Means with the same letter are not significantly different.			
Tukey Grouping	**Mean**	**N**	**diet**
A	1.11000	5	1
B	0.99200	5	2
C	0.43000	5	3

(d) Writing the additive model for block and diet effects as $Y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}$ for i=1,2,3 diets and j=1,2,3,4,5 blocks (age groups), two orthogonal contrasts are $\gamma_1 = (-1)\alpha_1 + (0)\alpha_2 + (1)\alpha_3$, a linear trend in the diet means, and , $\gamma_2 = (-1)\alpha_1 + (2)\alpha_2 + (-1)\alpha_3$ a quadratic trend in the diet means. This assumes the fat contents of the diets are nearly equally spaced. Contrast estimates, sum of squares, and values of F-tests and p-values for testing the null hypotheses that contrasts are zero are show below:

Contrast	Estimate	Contrast SS	Mean Square	F Value	Pr>F
liner trend	-0.680	1.15600000	1.15600000	478.67	<.0001
quadratic trend	0.444	0.16428000	0.16428000	68.02	<.0001

Both the linear and quadratic trends are significant.

(e)

The normal probability (Figure 1) shows the type of non-linear pattern that results from a distribution with a density function that has thicker right and left tails (more spread

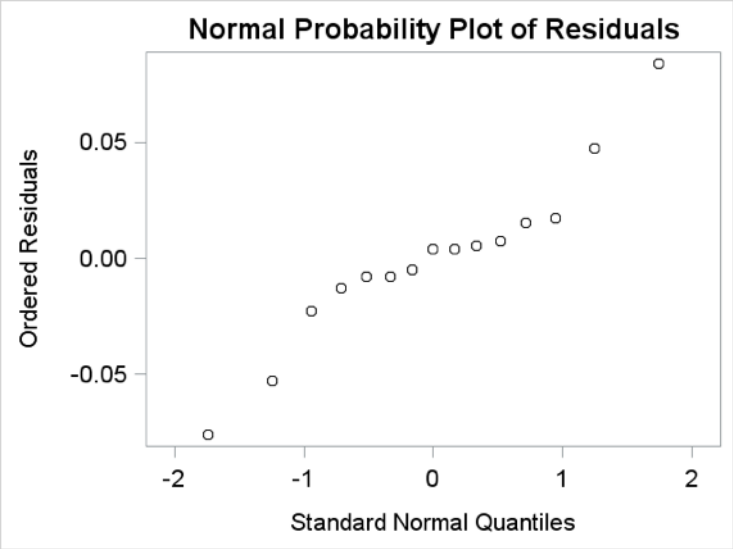


Figure 1: Q-Q plot.

out) than a normal distribution. The following tests, however, do not reject the null hypothesis of normally distributed random errors at the .05 level.

Tests for Normality				
Test	**Statistic**		**p Value**	
Shapiro-Wilk	**W**	0.93406	**Pr < W**	0.3135
Kolmogorov-Smirnov	**D**	0.187061	**Pr > D**	>0.1500
Cramer-von Mises	**W-Sq**	0.109059	**Pr > W-Sq**	0.0802
Anderson-Darling	**A-Sq**	0.577578	**Pr > A-Sq**	0.1134

(f) $\sigma^2_{RCBD} = MS_{error} = 0.002$.

$$\begin{aligned} & \sigma^2_{RCD} \\ = & \frac{(n-1)MS_{blocks} + n(J-1)MS_{error}}{nJ-1} \\ = & \frac{4 \cdot 0.355 + 5 \cdot 2 \cdot 0.002}{14} \\ = & 0.103 \end{aligned}$$

estimated efficiency $= \frac{(df_{RCBD}+1)(df_{RCD}+3)\sigma^2_{CRD}}{(df_{RCBD}+3)(df_{RCD}+1)\sigma^2_{CRD}} = \frac{(8+1)(12+3)0.103}{(8+3)(12+1)0.002} = 42.69$. To have the same efficiency, $n_{CRD} = 42.69 n_{RCBD}$.