Practice Questions for Exam #1 (with solutions)

1. qualitative: (c) (d) (g)

   discrete: (a) (c) (d) (f) (g)

2. tabular: Using a cross-tabulation of both variables.

   Graphical: multiple bar graphs. For example, for each party show three bar graphs: one corresponding to parents and child affiliated to that party, another to parents affiliated but child not, and the third for child affiliated but parent not.

3. You can compute the mean and the standard deviation to give an idea of the variation in the distribution.

4. $E[X] = \Sigma xP(x) = -2*.4 + 1*.6 = -.2$

   $Var(X) = \Sigma(x-E[X])^2 = (-2+.2)^2*.4 + (1+.2)^2*.6 = 2.16$

5. $Y = X*12$

   $\sigma_x = 0.6$ $Var(X) = 0.6^2 = .36$

$Var(Y) = 12^2 * Var(X) = 144*.36 = 51.84$ and $\sigma_y = 7.2$

6. The higher the national debt of a country the lower its GDP is likely to be. Lower national debts are associated with higher GDP.

7. N = 1000

p = .11

Var(p) = p*(1-p)/N = (.11*.89)/1000 = .0000979   s =.0099

$[p - Z_{\alpha/2} * s \; p + Z_{\alpha/2} * s]$ = [.11-1.96*.0099, .11+1.96*.0099] = [.09,.13]

8.

(a) Bar graphs for each category.

(b) The median is the category corresponding to “unfair”. We know that the first half of the sample provided answers that varied from “very fair” to “unfair”, while the second half finds redistribution in Brazil to be unfair or very unfair. The mode is also the “unfair” answer. From all possible answers, respondents were more likely to judge the income distribution in Brazil as unfair.

9.

(a) The mean is given by:

$\Sigma(x)P(x)$ = 0*0.06 + 1*0.03 + 2*0.05 + 3*0.08 + 4*0.13 + 5*0.35 + 6*0.09 + 7*0.07 + 8*0.06 + 9*0.03 + 10*0.05

= 4.92 for Mexico, and

$\Sigma(x)P(x) =$ 7.09 for Colombia

To get the standard error of our estimate of the mean, we first need to calculate the sample standard deviation:

$[\Sigma(x - E[x])^2 P(x)]^{1/2} = [(0-4.92)^2*0.06 + (1-4.92)^2*0.03 + (2-4.92)^2*0.05 + (3-4.92)^2*0.08 + (4-4.92)^2*0.13 + (5-4.92)^2*0.35 + (6-4.92)^2*0.09 + (7-4.92)^2*0.07 + (8-4.92)^2*0.06 + (9-4.92)^2*0.03 + (10-4.92)^2*0.05]^{1/2}$

$=$ 2.35 for Mexico, and

$[\Sigma(x - E[x])^2 P(x)]^{1/2} =$ 2.85 for Colombia

Based on the Central Limit Theorem we know the estimated mean is distributed approximately normal with variance σ^2/N and standard error $\sigma/\sqrt{N}$, where σ and N refer to the sample standard deviation and number of observations, respectively. Thus the standard errors are:

$2.35/\sqrt{927} = 0.08$ for Mexico, and

$2.85/\sqrt{716} = 0.11$ for Colombia

(b) Mexicans seem to embrace, on average, a more leftist ideological position than their Colombians counterparts. In order to know if this difference is statistically significant, we run a comparison of means test, as follows:

H_0: $\bar{x} - \bar{y} = 0$

H_A: $\bar{x} - \bar{y} \neq 0$

where *x* corresponds to the Colombian sample and *y* to the Mexican one.

$Z = (\bar{x} - \bar{y} - \mu_x - \mu_y)/(\sigma_x/\sqrt{N_x} + \sigma_y/\sqrt{N_y})$

$= (7.09 - 4.92)/(0.11 + 0.08) = 11.81$

$P(Z \geq 11.81)$ or $P(Z \leq -11.81) = 0.000$

Therefore, there is a statistically significant difference at the 0.01, 0.05, and 0.10 levels. We reject the null that citizens of both countries have the same ideological position. Colombians are significantly more “rightist” than Mexicans.

(c) In order to form a 95% confidence interval we need to know the standard errors of the proportions we are interested in.

For the first category we have:

$Var(p_y) = p*(1-p)/N = .06(.94)/927 = .00006$ for Mexico, and

$Var(p_x) = .02(.98)/716 = .00003$ for Colombia

$\sqrt{Var(p_x - p_y)} = (.00003 + .00006)^{1/2} = .0094$

with confidence interval:

$[p_x-p_y - Z_{\alpha/2} * \sqrt{Var(p_x - p_y)},\ p_x-p_y + Z_{\alpha/2} * \sqrt{Var(p_x - p_y)}]$

[.02-.06-1.96*.0094, .04 + 1.96*.0094]

[-.058,-.022]

For the extreme right category:

$Var(p_y) = p*(1-p)/N = .05(.95)/927 = .00005$ for Mexico, and

$Var(p_x) = .34(.66)/716 = .0003$ for Colombia

$\sqrt{Var(p_x - p_y)} = (.0003 + .00005)^{1/2} = .019$

with confidence interval:

[.25, .33]

With respect to the extreme left category, both countries display low proportions, however, Colombia displays a statistically significantly lower proportion if compared to Mexico (the confidence interval around the difference between the two proportions does not include zero, so we reject the null that there is no difference between the two countries at the 5% level).

With respect to the extreme right category, Colombia displays a statistically significantly higher proportion of citizens assuming that position. In this case the difference between the two countries is higher too, between 25% and 33%, at the 5% level.

10. H_0: $\mu = 20{,}000$

H_A: $\mu \neq 20{,}000$

N =21 cars $\bar{X} = 18{,}700$ $\sigma = 8{,}600$

$Z = (\bar{X} - \mu)/ \sigma/\sqrt{N} = -.69$

$2*P(Z\geq .69) = .4885$

Given that .4885 > .05 we fail to reject the null that the car manufacturer is right.

11. We need to run a difference in means matched pair test.

$\bar{D} = 0.3$, $s_D = .75$, N=30

H_0: $\mu_x = \mu_y$

H_A: $\mu_x \geq \mu_y$

$Z = \bar{D} - (\mu_{x0} - \mu_{y0})/(s_D/\sqrt{N})$

$Z = .3/.75/\sqrt{30}$

$Z = 2.19$

The critical value for a one-sided test with $\alpha = 5\%$ is $Z_{0.05} = 1.645$. Since Z > 1.645 we reject the null that the test had no effect.

12. Here we employ a test of proportion.

H_0: π = .5

H_A: π ≥ .5

p = 0.7

Var(p) = [p (1-p) /N] = [(.7*.3) /10] = 0.021

Std. Error: $[p (1-p) /N]^{1/2} = 0.021^{1/2} = 0.145$

Z = (.7-.5)/0.145 = 1.38

We only find support for the claim that the person possesses ESP if we are willing to employ a one-sided test at the 10% level. In this case the critical value of Z is 1.28 < 1.38. If we employ a two-sided test at the 10% level we would find no evidence of ESP. The same is true for two-sided and one-sided tests at the 5% and 1% levels.

13. Here we need a test of population mean.

H_0: μ = .25

H_A: μ < .25

N = 17 , σ = .056

$Z = (.23 - .25)/ .056/\sqrt{17} = -1.47$

This corresponds to a p-value of 0.1409 for a two-tailed test and 0.0704 for a one-tailed test. Thus, employing a 5% level of confidence, we would fail to reject the null that the deli is right. Again, the only way we would reject the null and conclude that we are being cheated is if we were to employ a one-tailed test at the 10% level.

14. Here we need to run a difference of means test on two independent samples.

H_0: $\mu_x = \mu_y$

H_A: $\mu_x \neq \mu_y$

Female (x): $\bar{x} = 11$, $\sigma_x = 1$, $N_x = 11$

Male (y): $\bar{y} = 14$, $\sigma_y = 3$, $N_y = 16$

$Z = (\bar{x} - \bar{y}) / (\sigma_x / \sqrt{N_x} + \sigma_y / \sqrt{N_y})$

$Z = (11-14) / (1/\sqrt{11} + 3/\sqrt{16}) = -3.7$

Since $|Z| > 2.57$ (the critical value for a two-tailed test at the 1% level) we reject the null of no discrimination at any conventional level of confidence.

15. Here we need a difference in proportions test.

H_0: $\pi_X = \pi_Y$

H_A: $\pi_X \neq \pi_Y$

Female (x): $p_x = 445/(445+675) = 40\%$ $N_x = 445+675 = 1120$

Male (y): $p_y = 515/(515+641) = 45\%$, $N_y = 1156$

$Z = [(p_x - p_y) - (\pi_x - \pi_y)]/[p_x (1-p_x)/N_x + (p_y (1-p_y)/N_y]^{1/2}$

$Z = .05/(.24/1120 + .2475/1156)^{1/2} = -2.42$

The corresponding p-value for a two-tailed test is 0.0157

We would, then, reject the null that males and females high school students are equally likely to consume marijuana at the 5% level, but we would not reject the null at the 1% level.

The 95% confidence interval around the estimated difference is:

[-0.0906, -0.0094]

Since it does not include zero and is negative, we find strong evidence that female high school students are less likely than males to consume marijuana at the 5% level.

16. In this case we need a difference in means test on two independent samples.

H_0: $\mu_X = \mu_Y$

H_A: $\mu_X \neq \mu_Y$

Buffalo (x): $\bar{x} = 121$, $\sigma_x = 18$, $N_x = 717$

Rochester (y): $\bar{y} = 114$, $\sigma_y = 13$, $N_y = 834$

$$Z = (\bar{x} - \bar{y}) / (\sigma_x / \sqrt{N_x} + \sigma_y / \sqrt{N_y})$$

$$Z = (121 – 114)/(18/\sqrt{717} + 13/\sqrt{834})$$

$$Z = 6.24$$

With such a high Z score we reject the null that house prices in Rochester are the same as in Buffalo at any conventional level of confidence. Looking at the 95% confidence interval around the difference in means, we find that houses in Buffalo tend to be more expensive than houses in Rochester.

$[7 – 1.96*(\sigma_x / \sqrt{N_x} + \sigma_y / \sqrt{N_y}) , 7 + 1.96*(\sigma_x / \sqrt{N_x} + \sigma_y / \sqrt{N_y})]$

[4.8, 9.2]

17. This problem requires a difference in means test for matched-pairs

$\bar{D}$ = 3450 – 3200 = 250, s_D = 300 , N=312

First, we consider whether the plan had a positive effect (using a one-sided test).

We have,

H_0: $\mu_X - \mu_Y = 0$

H_A: $\mu_X - \mu_Y \neq 0$

$$Z = \frac{\bar{D} - (\mu_{X,0} - \mu_{Y,0})}{s_D / \sqrt{N}} = \frac{250}{300/\sqrt{312}}$$

Z = 14.72, which corresponds to a p-value of .000

We reject the null at all conventional levels of confidence (10%, 5% and 1%).

In order to asses whether the plan was cost effective, we consider the null and alternative,

H_0: $\mu_X - \mu_Y = 300$

H_A: $\mu_X - \mu_Y \geq 300$

$$Z = \frac{\bar{D} - (\mu_{X,0} - \mu_{Y,0})}{s_D / \sqrt{N}} = \frac{250 - 300}{300/\sqrt{312}} = -2.94$$

For a one sided test, our rejection rule is to reject if $Z > Z_\alpha$ where $Z_{0.05} = 1.645$.

Thus, we fail to reject the null hypothesis that the plan had no impact on revenue.

This question is especially tricky because the alternative hypothesis (that the plan had a positive impact on revenue) is inconsistent with the finding that the plan had a negative impact of revenue.

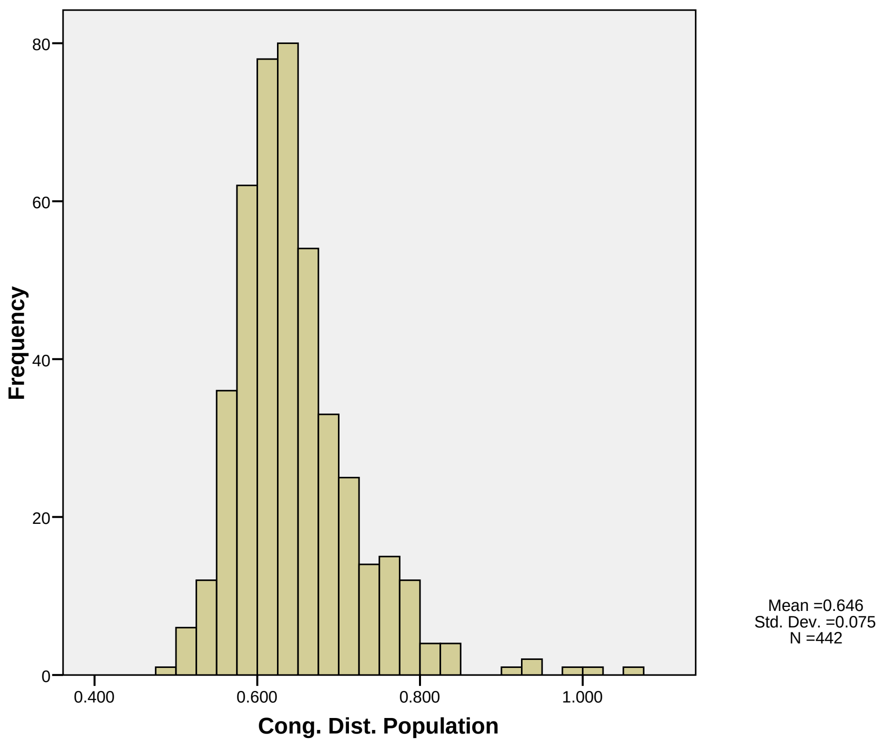
18.

(a) We can compute a measure of Skewness using the command Analyze → Descriptive Statistics → Descriptives, selecting options, and selecting Skewness. We obtain the result below,

	N	Minimum	Maximum	Mean	Std. Deviation	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
Cong. Dist. Population	442	.494	1.062	.64596	.074868	1.535	.116	4.603	.232
Valid N (listwise)	442								

The estimated skewness of 1.535 indicates that the data exhibit positive skewness, or that the data are skewed to the right.

(b) We can plot the distribution of district population using the Graphs → Legacy Dialogs → Histogram command. We obtain the following,



We can determine that the data are unimodal, since the above graph only has a single peak.

(c) Computing the correlation between these two variables, we have,

		Dem. House Cand. Ideology (NPAT)	Rep. House Cand. Ideology (NPAT)
Dem. House Cand. Ideology (NPAT)	Pearson Correlation	1	.304(**)
	Sig. (2-tailed)		.000
	N	355	299
Rep. House Cand. Ideology (NPAT)	Pearson Correlation	.304(**)	1
	Sig. (2-tailed)	.000	
	N	299	365

The correlation is 30.4%, indicating a moderately strong positive relationship. Districts that had Democratic candidates that were more conservative than average also tended to have Republican candidates that were more conservative than average.

(d) In this case, we will employ a matched pairs test. Let μ_D denote the population mean position of Democratic candidates and let μ_R denote the population mean position of republican candidates. We can state the null hypothesis as $H_0 : \mu_D = \mu_R$ and the alternative hypothesis as $H_A : \mu_D \neq \mu_R$. We have,

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Dem. House Cand. Ideology (NPAT)	.27324	299	.145636	.008422
	Rep. House Cand. Ideology (NPAT)	.74989	299	.144333	.008347

Paired Samples Test

		Paired Differences							
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t	df	Sig. (2-tailed)
Pair 1	Dem. House Cand. Ideology (NPAT) - Rep. House Cand. Ideology (NPAT)	-.476657	.171095	.009895	-.496129	-.457185	-48.173	298	.000

We can see that, as we expect, the population mean for the Republicans is greater than the population mean for the Democratic candidates. This conforms with our expectations that Republican candidates take more right-wing positions. We can also see by the p-value of 0.000 that the difference is significant at the 5% level.

(e) We can answer this question by employing an independent samples test.

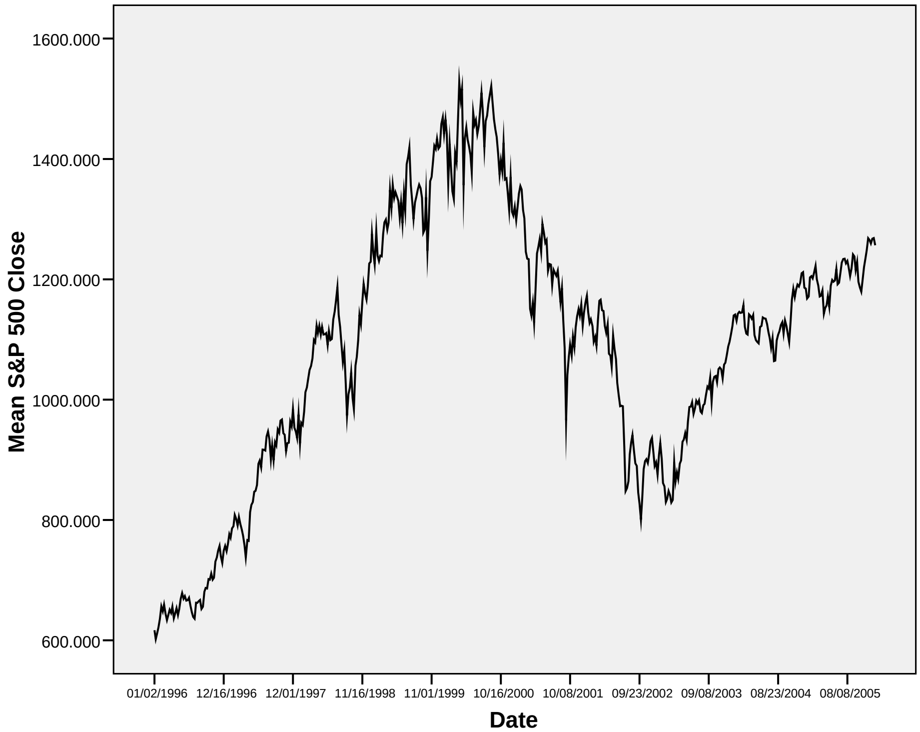
	Cong. Dist. in the South	N	Mean	Std. Deviation	Std. Error Mean
Cong. Dist. Ideology (Survey-based Measure)	0	284	3.10413	.182350	.010820
	1	155	3.27552	.144297	.011590

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Cong. Dist. Ideology (Survey-based Measure)	Equal variances assumed	7.528	.006	-10.101	437	.000	-.171395	.016968	-.204745	-.138046
	Equal variances not assumed			-10.809	381.663	.000	-.171395	.015856	-.202571	-.140219

Focusing on the second row (Equal variances not assumed), we find that the difference is statistically significant at the 5% level. Hence, the average congressional district in the south is more conservative than the average congressional district in the non-Southern United States.

19. A line graph of the time-series can be obtained by using the SPSS command, Graphs → Legacy Dialogs → Line, selecting ‘other statistic (e.g. mean), selecting S&P 500 close as the variable, and selecting ‘date’ as the category axes.



The results indicate that the S&P500 index was rising up until about the beginning of 2001, and then began to rapidly fall. Starting around 2003, the stock index began to rise again.

20.

(a) We can analyze this relationship using either a cross tabulation or computing the correlation coefficient. The correlation coefficient can be computed using Analyze → Correlate → Bivariate, and selecting the relevant variables.

		Return Golan to Syria	Right left
Return Golan to Syria	Pearson Correlation	1	.460(**)
	Sig. (2-tailed)		.000
	N	788	774
Right left	Pearson Correlation	.460(**)	1
	Sig. (2-tailed)	.000	
	N	774	1381

We can determine that the correlation is 46%. This is a moderately strong positive correlation, indicating that those with more left-wing ideologies are more likely to support returning the Golan Heights to Syria. This result conforms with our expectations- we expect views on security to dominate in the responses Israelis give to the left-right item, with right-wing views encompassing the beliefs that maintaining the occupied areas is essential for the security of Israel and with left-wing views encompassing the belief that returning some of the occupied areas are in Israelis long-term security interests, because they increase the likelihood of a peaceful settlement between Israel, the Palestinians, and neighboring Arab states.

Using a cross-tabulation, we have,

			Right left							Total
			Right	2	3	4	5	6	Left	Right
Return Golan to Syria	Give back none	Count	97	64	54	96	30	19	19	379
		% within Right left	78.9%	71.1%	65.9%	47.3%	28.6%	22.4%	22.1%	49.0%
	Give back small part	Count	20	21	17	33	27	10	18	146
		% within Right left	16.3%	23.3%	20.7%	16.3%	25.7%	11.8%	20.9%	18.9%
	Give back large part	Count	1	2	11	29	14	16	20	93
		% within Right left	.8%	2.2%	13.4%	14.3%	13.3%	18.8%	23.3%	12.0%
	Give back all	Count	5	3	0	45	34	40	29	156
		% within Right left	4.1%	3.3%	.0%	22.2%	32.4%	47.1%	33.7%	20.2%
Total		Count	123	90	82	203	105	85	86	774
		% within Right left	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

In this cross-tabulation, we see the same pattern. The vast majority of those holding very right wing views don't support giving any of the Golan Heights back to Syria, while those holding the most left-wing views are likely to support giving some or all of the Golan Heights back to Syria.

(b) Let μ_X denote the population mean of A12 and let μ_Y denote the population mean of A13. We can state the null hypothesis as $H_0 : \mu_X = \mu_Y$ and the alternative hypothesis as $H_A : \mu_X \neq \mu_Y$. We can test the null hypothesis in SPSS using the Analyze → Compare Means → Paired Samples T Test in SPSS. Selecting the variables, we have,

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Arab areas in Jerusalem to palestine	2.33	1390	.982	.026
	Temple mount to palestine	2.06	1390	.965	.026

		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference (Lower, Upper)		t	df	Sig. (2-tailed)
Pair 1	Arab areas in Jerusalem to palestine - Temple mount to palestine	.274	.835	.022	.230	.318	12.238	1389	.000

The results indicate that more Israelis support return the Arab areas than support returning the Temple Mount. The difference is statistically significant at the 0.1% level (since the p-value is zero to 3 decimal places).