Contents

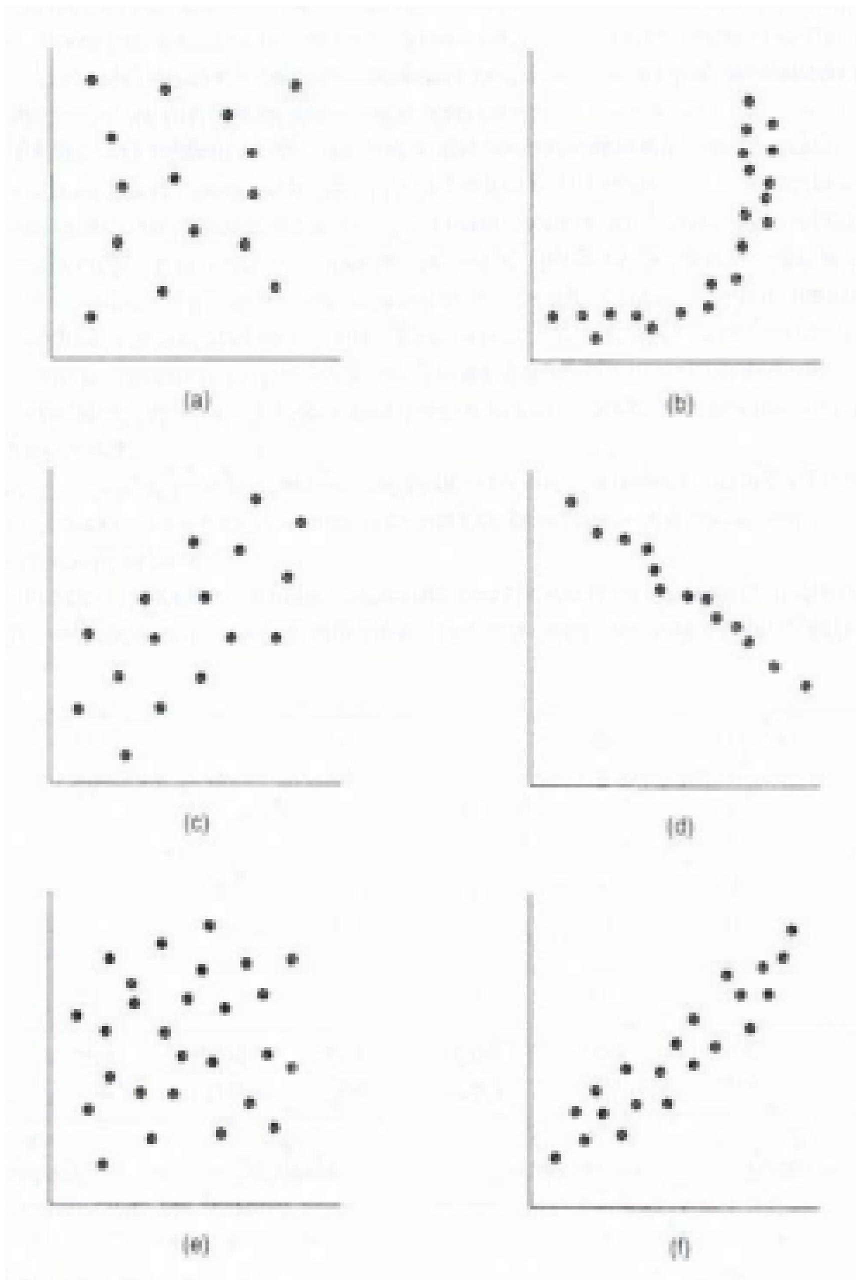
1 Notes

1.1 The Hypothesis-Testing Process

- Step 1: Restate the Question as a Research Hypothesis.
- Step 2: Determine the Characteristics of the Comparison Distribution.
- Step 3: Determine the Cutoff Sample Score on the Comparison Distribution at Which the Null Hypothesis Should Be Rejected.
- Step 4: Determine Your Sample's Score on the Comparison Distribution.
- Step 5: Decide Whether to Reject the Null Hypothesis.

2 Exercise 11

"For each of the following scatter diagrams, indicate whether the pattern is linear, curvilinear, or no correlation; if it is linear, indicate whether it is



2.1 Resolution

First, let's state the equation which defines correlation, by Pearson:

$$r_{xy} = \frac{\sum_{i=1}^{n} (x_i - \overline{x}).(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \overline{x})^2}.\sqrt{\sum_{i=1}^{n} (y_i - \overline{y})^2}} \tag{1}$$

Or, for a population,

$$\rho_{xy} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y} \tag{2}$$

2.1.1 Pattern (a)

If we were to fit the best line, or curve, there would be for each point an approximately an opposite point which would give us an error, regardless of the direction of the curve. Therefore, it's correlation pretty close to 0. Because, $E[(X - \mu_X)] \wedge E[(Y - \mu_Y)]$ both tends to zero.

The numerator reminder, being Equation 2, $E[(X - \mu_X)(Y - \mu_Y)] = \lim_{E(X) \to 0,\, E(Y) \to 0} (E[(XY)] - E[X]E[Y])$

Which gives us, $E[(X - \mu_X)(Y - \mu_Y)] = \lim_{E(X) \to 0,\, E(Y) \to 0} (E[(XY)])$.

$\therefore p \approx 0$.

2.1.2 Pattern (b)

Notation: A is a constant to be tuned.

A fitting-curve can be approximated to $y(x) = A.x^2$. This pattern is quadratic. Also, we are dealing with values $x > 0$ of our ideal-approximation of $y(x)$. Therefore, the correlation remains high, $\therefore p \to 1$ for curvilinear fitting.

2.1.3 Pattern (C)

For a given x_i there are approximately three y_j corresponding to it. At the same time, the y_j are spaced. But, the overall trend upwards is clear, implying p is positive. Therefore, there exist a week correlation, estimated about $0.3 < p < 0.7$.

2.1.4 Pattern (d)

For each x_i there is one clear y_i, and the fitted curve as $y(x) = A.x$, gives us a almost perfect fit. It's linear and $p \to -1$.

2.1.5 Pattern (e)

Likewise to pattern (a), $p \to 0$. But, the value would be greater than Pattern (a), because the experimental points are closer to each other.

2.1.6 Pattern (f)

The pattern is linear, and it's absolute value for correlation would be less than Pattern (d) and greater than (c). Also, as the trend is upwards, the value of p would be positive.

3 Exercise 12

As part of a larger study, Speed and Gangestad (1997) colle14cted ratings and nominations on a number of characteristics for 66 fraternity men from their fellow fraternity members. The following paragraph is taken from their *Results* section: "men's romantic popularity significantly correlated with several chacteristics: best dressed (r = .47), most self-confident (r = .48), best trend- setters (r = .38), funniest (r = .37), most satisfied (r = .32), and most independent (r=.28). Unexpectedly, however, men's potential for financial success did not significantly correlate with romantic popularity (r = .10)." (p.931)

Explain these results as if you were writing to a person who has never had a course in statistics. Specifically, (a) explain what is meant by a correlation coefficient using one of the correlations as an example; (b) explain in a general14 way what is meant by significant and not significant, referring to at least one specific example; and (c) speculate on the meaning of the pattern of results,

3.1 Resolution

3.1.1 12(a) - Correlation meaning

Using Pearson's definition Equation 1, the denominator tends to zero as the mean variable is closer to each data points, which makes $|r_{xy}|$ grow. That is, firstly, correlation measures how far apart data points are.

Secondly, it measures how much a trend can be modeled linearly. That is, the mean value always will be on the center of a straight line. Therefore, maximizing the even-spacement of each data and the value of r_{xy}.

Third, the value is not modularized. Thus r_{xy} can be both negative. It gives a sense it the trend of y grow as decreases as x increases. A fourth important component, but more technical one, is that the Pearson's value is limited above and below. The value being normalized, $|r_{xy}| < 1$, we can compare the meaning of correlation among different sets of measures. e.g., height, age, income, etc.

3.1.2 12(b) - Correlation Significance

Correlation significance gives a limited-number, for how predictably linear data-behavior is. So, taking the extremes. If $r_{xy} = 0$ means there is no way of telling how an increase in x will result in a increase of decrease of y, while modeling the data-behavior by a straight-line. At the other hand, $r_{xy} = 1$ means data behaves exactly as a straight-line. So, an increase in Δx means an increase in $\Delta y = A.\Delta x$ - it's totally certain that this is to be observed for any Δx.

3.1.3 12 (c) - Speculation about meaning

The traits of "good-fitness" to the environment correlates positively with female-selection of partners. e.g., well dressed, self-confident, etc. This trend is linear, but mid-strong ($0.3 < r < 0.5$). At the same time, yet another heavy measure of "good-fitness" to environment - money -, does not strongly correlates to being selected by females.

It could be argued that the social-genetic peer-selection factors haven't had enough time to catch the human trend of using money as a mean of attaining high standards of living; but that would be to doubt human intelligence. The other plausible explanation could be that **income** being the ultimate social "good-fitness" measure to environment, is not linearly correlated to female-selection. Therefore, an increase in income by double increases the chance of female-selection by much more than the double, and this rate changes as higher the income is.

4 Exercise 13

Gable and Lutz (2000) studied 65 children, 3 to 10 years old, and their parents. One of their results was: "Parental control of child eating showed a negative association with children's participation in extracurricular activities ($r = -.34; p < .01$)." Another result as Parents who held less appropriate-beliefs about children's nutrition reported that their children watched more hours of television per day (r = .36; p<.01)." (Both quotes from page 296.*** 14Explain these results as if you were writing to a person who has never had a course in statistics. Be sure to comment on possible directions of causality for each result.

4.1 Resolution

4.1.1 Taking the researches separately

These two trends have meaning, wh14en analyzed separately. At the same time, an expeculative meaning can be derived analyzing them together.

We should note that with a high degree of certainty on reproducibility,there exist a trend in both data. e.i., $p < .01$.

The correlation established says:

- As more control of child eating is exerted, the lesser chance of the chil***p14articipating in extracurricular activity.
- Also, the less knowledgeable and careful of child nutrition a parent is,

4.1.2 Taking them together

Together, some hypothesis can be formulated. First, that control of child eating can lead to the child spending less time on television and therefore14 more time on learning. As a consequence, not needing or feeling the need of***articipating in extracurricular activity.

Other hypothesis could be that children who watch more television is more prone to participate in extracurricular activity, due to being a greater student - which is contra-intuitive to the notion that television's programsare mostly useless.

More research should be done to explain the relation between child eating control, extracurricular activity, television time use, and parent***n14otions around eating.

5 Exercise 14

"Suppose you want to conduct a survey of the attitude of psychology graduate students studying clinical psychology towards Freudian methods of psychotherapy. One approach would be to contact every psychology graduate student you know and ask them to fill out a questionnaire about. (a) What kind of sampling method is this? (b) What is a major limitation of this kind of approach?"

5.1 Resolution

5.1.1 14 (a) Sampling method

It would be a cluster sample. The researcher would be using his social-network as the representative population of all graduate students studying clinical psychology.

5.1.2 14 (b) Limitation

That if these particular students are biased in some way, the conclusions could not be applied to any other random population of students. Thus, conclusions may only apply to this sample - there lacks variability in sample.

6 Exercise 15

"A large study of how people make future plans and its relation to their life satisfaction (Prenda & Clachman, 2001) obtained their participants through random-digit dialing procedures These are procedures in which phone numbers to call potential participants are selected at random from all phone numbers in a particular country. Explain to a person who has never had a course in statistics (a) why this method of sampling might be used, and (b) why it may be a problem if not everyone called agreed to be interviewed."

6.1 Resolution

6.1.1 15 (a) Purpose of this sampling method

By selecting people at random, is the hole population of a country, there is no implicit pattern that could be distinguished for each person. Therefore, the results are much more suited to generalization. That is, if the results showed a determined trend, it would be safe to say the trend applies to the rest of the population - it's a representative relation.

6.1.2 15 (b) Problem on certain people not answering the phone

The data may not be suited to generalization, then. Because, this group of people could have some common trait of behavior that could effect significantly the measure of a trend on the capacity to "make future plans". In general, it would be doubtful if the conclusions arrived applied also to the category of people who refused to take random calls of strangers.

7 Exercise 16

"Suppose that you were going to conduct a survey of visitors to your campus. You want the survey to be as representative as possible. How would you select the people to survey? Why would that be your best method?"

7.1 Resolution

The first thing was to find records on people who visited the campus across time. If such record existed, the next step would be to randomly select the individuals of this list. This way, there would be no bias on the research regarding the time these people visited the university; what are their age; or the frequency of their visits.

8 Exercise 17

"Define the following terms in your own words: (a) hypothesis-testing procedure, (b) .05 significance level, and (c) two-tailed test."

8.1 Resolution

8.1.1 17(a) Hypothesis-testing

Hypothesis-testing procedure means the entire process from selecting the method to selecting samples, to that of analyzing the data and arriving on a conclusion, with a degree of certainty of how representative that hypothesis of the general population.

8.1.2 17(b) .05 significance level

A .05 significance level would be an estimate of how likely that result is among samples. So, in this case, out of 100 random samples, we would hope that 95 of these would follow the conclusions arrived. And, 5 of them could

8.1.3 17(c) two-tailed test

The two-tailed test can show us if a certain relation, concerning an hypothesis, can be of used to understand some aspect of a population or not. If the test fails, this means there is some relation among how two variables behave themselves, inside the population.

9 Exercise 18

"List five steps of hypothesis testing and explain the procedure and logic of each."

9.1 Resolution

- Formulating an hypothesis: review the literature on a subject, and for-

mulate explanations to previously non-explained behavior. Or, choose an hypothesis verified in the literature, so to verify if it's possible to replicate the results. State this in mathematical terms.

- Choosing a variable to measure: determine what variable could help

inquiry into the hypothesis.

- Determine the Cutoff Sample Score on the Comparison Distribution at Which the Null Hypothesis Should Be Rejected: determine a sampling method that will dictate how representative your conclusion is. Also, We should *a priori* know how our cutoff will be done (level of significance).
- Determine Your Sample's Score on the Comparison Distribution: quantify our data; use the stablished frame of reference to test our data. We have to compute where the actual values lay, regarding cutoffs.
- Decide Whether to Reject the Null Hypothesis: based on the score, awnser our initial question of the validity of the reaseach hypothesis.

10 Exercise 19

"When a result is significant, explain why is it wrong to say the result proves the research hypothesis?"

The term proof is used, in the mathematical sense, that some relation **always** holds true. But, in statistics, we measure likelihoods, which imply there could be a sample that is so extreme that it falls under a condition that the hypothesis do not apply. That is, to use the hypothesis in any populations, as true, could turn out to actually be false for some population.

11 Exercise 20

"For each of the following:

1. say what two populations are being compared,
2. state the research hypothesis,
3. state the null hypothesis, and
4. say whether you should use a one-tailed or two-tailed test and why.

   - In an experiment, people are told to solve a problem by focusing on the details. Is the speed of solving the problem different for people who get such instructions compared to people who are given no special instructions?
   - Based on anthropological reports in which the status of women is scored on a 10-point scale, the mean and standard deviation across many cultures are known. A new culture is found in which there is an unusual family arrangement. The status of women is also related in this culture. Do cultures with the unusual family arrangement provide higher status to women than cultures in general?"

11.1 Solution

11.1.1 1.1 Told to concentrate

1. Populations The populations are:
   - People who have been told to focus and solve the problem;
   - People who haven't been told to focus, only to solve the problem;
2. Hypothesis The mean time to solve a problem, when the subjects are told to focus is different from when they are not.
3. Null Hypothesis The mean time of both populations are the same.
4. One or two tailed A two-tailed test is suited. Because we don't suspect if the mean time - if different - will be higher or lower, when the subjects are told to concentrate.

11.1.2 1.2 Status of women

1. Populations The populations are:
    - The general population of families across different cultures.
    - The population with odd family arrangements.
2. Hypothesis The mean status of women in a odd-arrangement family are is higher than found on the general orthodox family arrangements.
3. Null Hypothesis There is no difference in women's mean status between orthodox and odd-arranged families.
4. One or two tailed One tailed is best suited. Because we want to know about a predetermined difference in observable values. That is, we expect that one of them is greater than the other.

12 Exercise 21

"A researcher predicts that listening to music while solving math problems will make a particular brain area more active. To test this, a research participant has her brain scanned while listening to music and solving math problems, and the brain area of interest has a percent signal change of 58. From many previous studies with the same math-problems procedure (but not listening to music), it is known that the signal change in this brain area is normally distributed with a mean of 35 and a standard deviation of 10. Using the .01 level, what should the researcher conclude? Solve this problem explicitly using all five steps of hypothesis testing and illustrate your answer with a sketch showing the comparison distribution, the cutoff (cutoff), and the score of the sample on this distribution. Then explain your answer to someone who has never had a course in statistics (but who is familiar with mean, standard deviation, and Z scores)."

12.1 Solution

12.1.1 Population used

We have the distribution data of people who solved math problems and their increase in brain activity. But not of people who was both listening music and resolving math problems. Therefore, we will use this population as out control-population.

12.1.2 State the research hypothesis

The increase in the value of increased brain-activity found in the study-case is relevantly higher than the mean value of increase of people solving math problems, without music ($\alpha = 0.01$).

12.1.3 State the null hypothesis

The value observed is not as significant. That is, $H_0 : p > \alpha$

12.1.4 Computing the p-value

Using Python's statistics scientific library scipy.stats,

```
from scipy.stats import norm
```

The cumulative distribution function (CDF), given by

$$\text{norm.cdf}(x, \mu, \sigma) = \int_{-\infty}^{x} \frac{e^{-\frac{1}{2}\left(\frac{\xi-\mu}{\sigma}\right)^2}}{\sigma\sqrt{2\pi}} \tag{3}$$

Which means the area under the curve until the value x, the observable variable under comparison.

Therefore, the p-value is the $p_{value} = 1 - \text{norm.cdf}$

```
print(1 - norm.cdf(45, 35, 10))
```

RESULTS: 0.15865525393145707

12.1.5 Conclusion

$\therefore p_{value} > 0.01$, we don't reject H_0. The value found do not support the thesis that listening music further increases the activation of the brain activity in the determined area any more than just resolving mathematical problems.

12.1.6 Graphic

1. Find cutoff Finding x such that $p(x) = \alpha$,

   ```
   print(1 - norm.cdf(58.2633, 35, 10))
   ```

   RESULTS: 0.010000476391382573

   Therefore, the **cutoff** is at $x = 58.2633$.

2. Plot with cutoff and the presented value under the Normal

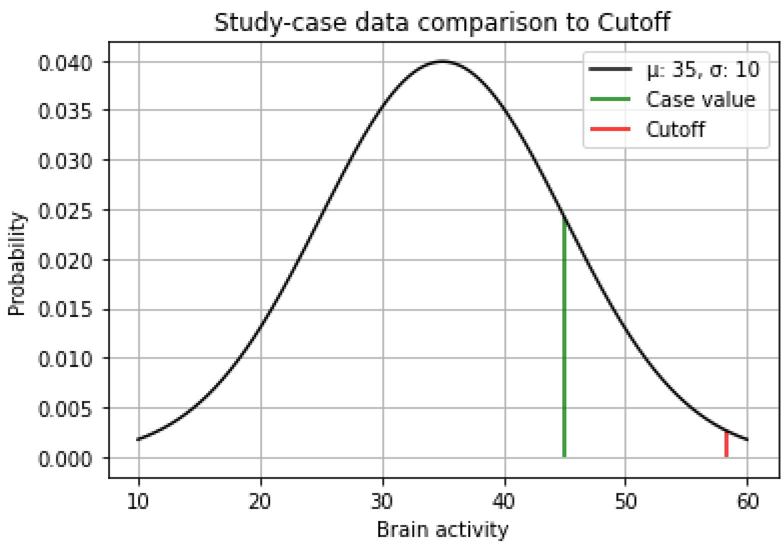
```
import numpy as np
import matplotlib.pyplot as plt

#x-axis ranges from 20 and 50 with .001 steps
x = np.arange(10, 60, 0.001)

#define normal values (not normalized)
plt.plot(x, norm.pdf(x, 35, 10), label=': 35, : 10', color='k')
plt.vlines(x = 45.0, ymin=0, ymax=norm.pdf(45, 35, 10), colors = 'green', label =
plt.vlines(x = 58.2633, ymin=0, ymax=norm.pdf(58.2633, 35, 10), colors = 'red', la

# Grid on
plt.grid(True)
# Title
plt.title('Study-case data comparison to Cutoff')
# Axis titles
plt.xlabel('Brain activity')
plt.ylabel('Probability')
#add legend to plot
plt.legend()
```

RESULTS:



13 Exercise 22

"In an article about anti-tobacco campaigns, Siegel and Biener (1997) discuss the results of a survey of tobacco usage and attitudes, conducted in Massachusetts in 1993 and 1995; Table 6-2 shows the results of this survey. Focusing on just the first line (the percentage smoking >25 cigarettes daily), explain what this result means to a person who has never had a course in statistics. (Focus on the meaning of this result in terms of the general logic of hypothesis testing and statistical significance."

TABLE 6–2 Selected Indicators of Change in Tobacco Use, ETS Exposure, and Public Attitudes Toward Tobacco Control Policies—Massachusetts, 1993–1995

	1993	1995
Adult Smoking Behavior		
Percentage smoking >25 cigarettes daily	24	10*
Percentage smoking <15 cigarettes daily	31	49*
Percentage smoking within 30 minutes of waking	54	41
Environmental Tobacco Smoke Exposure		
Percentage of workers reporting a smokefree worksite	53	65*
Mean hours of ETS exposure at work during prior week	4.2	2.3*
Percentage of homes in which smoking is banned	41	51*
Attitudes Toward Tobacco Control Policies		
Percentage supporting further increase in tax on tobacco with funds earmarked for tobacco control	78	81
Percentage believing ETS is harmful	90	84
Percentage supporting ban on vending machines	54	64*
Percentage supporting ban on support of sports and cultural events by tobacco companies	59	53*

Source: Biener and Roman. 1996.

*$p < .05$

Note. Data from Siegel, M., & Biener, L. (1997), tab. 4. Evaluating the impact of statewide anti-tobacco campaigns: The Massachusetts and California tobacco control programs. *Journal of Social Issues, 53*, 147–168. Copyright © 1997 by the Society for the Psychological Study of Social Issues. Reprinted with permission.

13.1 Solution

Twenty five cigarettes are not as extreme a value as to say it would be rare to find a people on the general population that smoked more cigarettes than that daily. To consider a value to be rare to be found in a population, the percentage value have to be lesser than 5%.

14 Exercise 23

"Define alpha and beta."

14.1 Solution

Alpha decreases, when we increase the Confidence Level we want to scrutinize our test. So, the lower we set Alphas, the harder is to have false positives (type I error) passing our test.

Beta is how much, percentage wise, we would accept to wrongly categorize data that collaborate to the alternative hypothesis. That is, how much false negatives (type II error) we are willing to commit, so to preserve our null hypothesis.

15 Exercise 24

"In a planned study, there is a known population with a normal distribution, $\mu = 15$, $\sigma = 2$. What is the predicted mean if researchers predict

- A small positive effect size,
- A medium negative effect size,
- a large positive effect size,
- An effect of d = .35, and
- An effect size of d = -1.5?"

15.1 Solution

15.1.1 Mathematical translation of terms

Cohen and Sawilosky's suggestions are of

Effect size	d
Very Small	0.01
Small	0.20
Medium	0.50
Large	0.80
Very large	1.20
Huge	2.0

The general d-value is computed as $d = \dfrac{|\overline{x_1} - \overline{x_2}|}{s}$, for which we will use the values of $x_1 = 15 \wedge s = 2$.

Generally, we have $\overline{x_2} = \overline{x_1} \pm s.(d)$

15.1.2 Predicting the means

Therefore, consulting the table for the d values,

1. Small positive effect size $\implies \overline{x_2} = 15.4$.
2. Medium negative effect size $\implies \overline{x_2} = \overline{x_1} - s.(d) \Leftrightarrow \overline{x_2} = 14$
3. A large positive effect size $\implies \overline{x_2} = \overline{x_1} + s.(d) \Leftrightarrow \overline{x_2} = 16.6$
4. An effect of d=.35, $\implies \overline{x_2} = \overline{x_1} + s.(d) \Leftrightarrow \overline{x_2} = 15.70$
5. An effect of d=-1.5 $\implies \overline{x_2} = \overline{x_1} - s.(d) \Leftrightarrow \overline{x_2} = 12$

16 Exercise 25

"Based on a particular theory of creativity, a psychologist predicts that artists will be greater risk takers than the general population. The general population is normally distributed with a mean of 50 and a standard deviation of 12 on the risk-taking questionnaire this psychologist plans to use. The psychologist expects that artists will score, on the average, 55 on this questionnaire. The psychologist plans to study 36 artists and test the hypothesis at the .05 level.

- What is the power of this study?
- Sketch the distributions involved, showing the area for alpha, beta, and power.
- Explain your answer to someone who understands hypothesis testing with means of samples but has

never learned about power."

16.1 Solution

We will be using statsmodels.stats.power.TTestIndPower, a library por Power Analysis in Python.

```
import statsmodels.stats.power as tt
```

But, under the hood we are considering a Z-statistic, of the following form:

$$Z = \frac{\overline{X} - \mu_1}{\left(\frac{\sigma}{\sqrt{n}}\right)} \tag{4}$$

In which,

$$\begin{cases} \overline{X} & : \text{The mean proposed in the hypothesis} \\ \mu_1 & : \text{The given mean} \\ \sigma & : \text{The given deviation} \\ \sqrt{n} & : \text{The size of the tested population} \end{cases} \tag{5}$$

16.1.1 What is the power of this study?

Let effect-size be our already defined d.

```
# difference in means divided by the standard deviation
effect_size = (55 - 50)/12
# number of observations
n_obs = 36
# alternative hypothesis: larger mean
alt = 'larger'
# alpha: .05 level
alpha= 0.05

## Calling the solver for power
tt.tt_ind_solve_power(effect_size=effect_size, nobs1=n_obs, alternative=alt, alpha=alpl
```

RESULTS: The power is of 54.

16.1.2 Sketch the distributions involved, showing the area for alpha, beta, and power.

```
from scipy.stats import norm
```

```
import numpy as np
import matplotlib.pyplot as plt
```

1. Finding the cutoff (try and error)

   ```
   print(1 - norm.cdf(69.7, 50, 12))
   ```

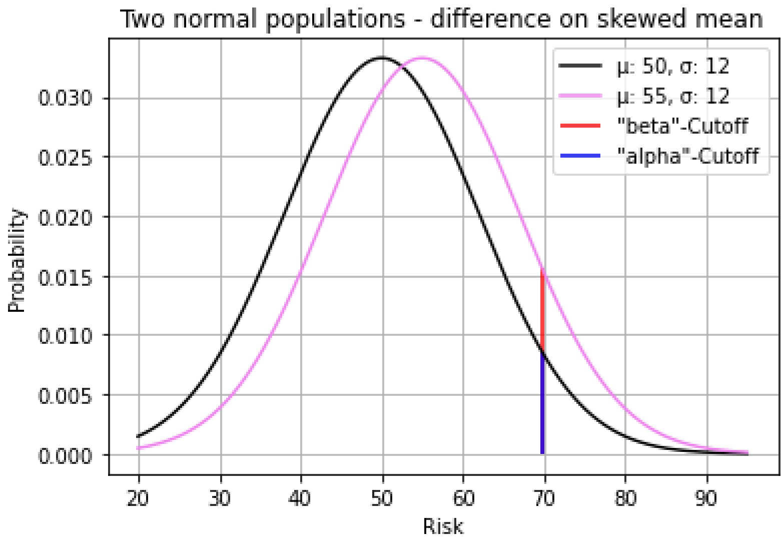
   Results: 0.05032955164661035

2. Sketch two distributions

   ```
   #x-axis
   x = np.arange(20, 95, 0.001)

   #define normal values (not normalized)
   plt.plot(x, norm.pdf(x, 50, 12), label=': 50, : 12', color='k')
   plt.plot(x, norm.pdf(x, 55, 12), label=': 55, : 12', color='violet')
   plt.vlines(x = 69.7, ymin=0, ymax=norm.pdf(69.7, 55, 12), colors = 'red', label =
   plt.vlines(x = 69.7, ymin=0, ymax=norm.pdf(69.7, 50, 12), colors = 'blue', label =

   # # Grid on
   plt.grid(True)
   # Title
   plt.title('Two normal populations - difference on skewed mean')
   # Axis titles
   plt.xlabel('Risk')
   plt.ylabel('Probability')
   #add legend to plot
   plt.legend()
   ```



- The area under the blue - alpha-cutoff - of the $N(\mu = 50, \sigma = 12)$, from x=69.7 to positive-infinity will be the I-type error commited (α).
- The area under the red - beta-cutoff - of the $N(\mu = 55, \sigma = 12)$, from x=69.7 to negative-infinity will be the II-type error commited (β).

3. Sketch Z distribution and Power

   Calculating z such P(Z>z)=1-P(Z<z)=0.54. That is, P(Z<z)=0.46

```
norm.cdf(4.80, 5, (12/np.sqrt(36)))
```

   RESULTS: 0.46

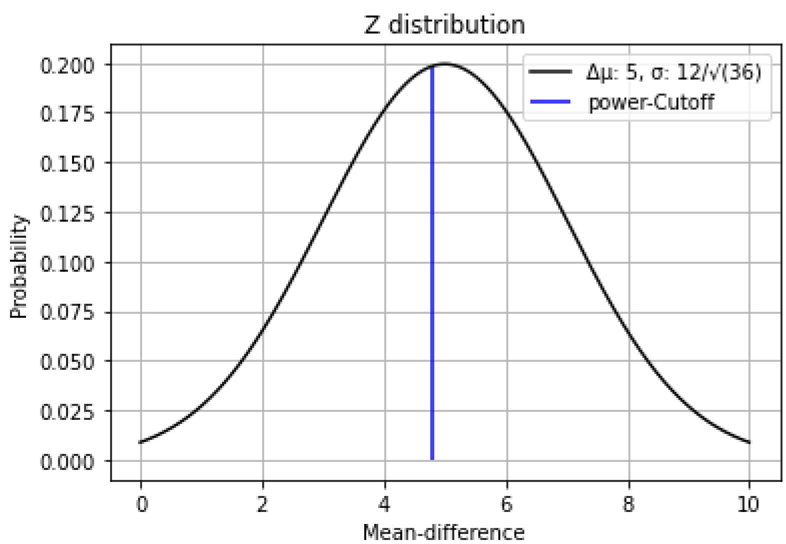
   So, z=0.46. This will be used to plot the cutoff region for understanding Power.

```
#x-axis
x = np.arange(0, 10, 0.001)

#define normal values (not normalized)
plt.plot(x, norm.pdf(x, 5, (12/np.sqrt(36))), label=': 5, : 12/(36)', color='k')
plt.vlines(x = 4.80, ymin=0, ymax=norm.pdf(4.80, 5, (12/np.sqrt(36))), colors = 'b
```

```
# # Grid on
plt.grid(True)
# Title
plt.title('Z distribution')
# Axis titles
plt.xlabel('Mean-difference')
plt.ylabel('Probability')
#add legend to plot
plt.legend()
```

```
<matplotlib.legend.Legend at 0x7fb975035e80>
```



The area above the power-Cutoff line adds up to 0.54. This is the power of our data.

16.1.3 Explain your answer to someone who knows Hypothesis testing (. . .)

We would need to increase the number of artists test to make the results relevante. This in case if they indeed follow the hypothesized increase in mean.

As the data is presented, we would be having a likelehood of 46% of misinterpreting out data. That is, if we ended up rejecting H_0, only 56%

of those deviant data would, for certain, be due to a skewed-distribution-behavior.

17 Exercise 26

"You read a study that just barely fails to be significant at the .05 level. That is, the result is not significant. You then look at the size of the sample. If the sample is very large (rather than very small), how should this affect your interpretation of:

- The probability that the null hypothesis is actually true, and
- The probability that the null hypothesis is actually false?"

17.1 Solution

If the sample is very large, the Center Limit Theorem (CLT) says the distribution should become a normal bell-curve.

17.1.1 The probability H_0 actually true

Then, this barely measurable difference won't change, when we increase even more the sample size. The result will very likely stay in the "not significant" category. Low probability of being false.

17.1.2 The probability H_0 actually false

Just the opposite of the true case, the likelihood that

18 Exercise 27

"You are planning a study that you compute as having quite low power. Name five things that you might do to increase power."

18.1 Solution

- Number of experiments (n).
- Change the effect size under hypothesis.
- Reframe the hypothesis to other values (μ).

- Compute the values to different α.
- Change the alternative hypothesis.

19 Exercise 28

"Evolutionary theories often emphasize that humans have adapted to their physical environment. One such theory hypothesizes that people should spontaneously follow a 24-hour cycle of sleeping and waking even if they are not exposed to the usual pattern of sunlight. To test this notion, eight paid volunteers were placed (individually) in a room in which there was no light from the outside and no clocks or other indications of time. They could turn the lights on and off as they wished. After a month in the room, each individual tended to develop a steady cycle. Their cycles at the end of the study were as follows: 25, 27, 25, 23, 24, 25, 26, and 25.

Using the 5% level of significance, what should we conclude about the theory that 24 hours is a natural cycle? (That is, does the average cycle length under these conditions differ significantly from 24 hours?)

- Use the steps of hypothesis testing.
- Sketch the distributions involved.
- Explain your answers to someone who has never taken a course in statistics."

19.1 Solution

19.1.1 28(a) Hypothesis testing steps

- Step 1: Restate the Question as a Research Hypothesis.
- Step 2: Determine the Characteristics of the Comparison Distribution.
- Step 3: Determine the Cutoff Sample Score on the Comparison Distribution at Which the Null Hypothesis Should Be Rejected.
- Step 4: Determine Your Sample's Score on the Comparison Distribution.
- Step 5: Decide Whether to Reject the Null Hypothesis.

1. S1: Nondirectional The mean day-time cycle found in people devoided of light and timers is not the same as the the general population. $H_0 : X - 25 = 0 \quad H_1 : X - 25 \neq 0$.

2. S2: p-values and t-statistic

   (a) Theoretical discussion We will be using a One-Sample T-Test. Because there is few measures of a sample, and we have a hypothesized mean to compare.

   So,

   - $\mu_0 = 24.0$
   - $M = 25.0$
   - $\sigma_1 = 1.2$
   - $n = 7$

   We could derive all relevant factors by hand, in this case:

   - $S^2_{\text{pooled}} = \dfrac{\sigma^2}{n-1}$
   - $S_M = S_{\text{pooled}}$
   - $t = \dfrac{M - \mu_0}{S_M}$

   But, we will use the already implemented libraries, so there is no mistake.

   (b) Numerical libraries Using the numerical library and the scientific libraries in Python,

```
import numpy as np
import scipy.stats as st
```

   Thankfully, Python has an already implemented modulus to test this kind of hypothesis

```
# State the data
data = [25,27,25,23,24,25,26]
# The hypothesized population mean - 24 hours cycles
popmean = 24.0

tStat, pValue =  st.ttest_1samp(data, popmean, axis=0)
print("P-Value:{0:.2f} T-Statistic:{1:.2f}".format(pValue,tStat)) #print the 

#       print("mean: {0:.2f}; standard deviation: {1:.2f}; S_pooled= {2:.2f}"
```

RESULTS: P-Value: 0.09 T-Statistic: 2.05

3. S3: Cutoff We will test mean differences for different populations. Also, we'll use $\alpha = 0.05$. As it's nondirectional, $\alpha_{\pm\frac{1}{2}} = 0.025$.

4. S4: Comparison to our data We had that the p-value = 0.09. As $p_{\text{value}} > \frac{\alpha}{2}$ (nondirectional) we fail to reject H_0. We could also consult a table with $t_{0.975,7} = 2.365$. As the actual t, $t < t_{0.975,7}$ then we don't reject H_0.

5. S5: Conclusion So, the means of the populations are, in fact, the same, under our test and scrutiny. There is evidence to collaborate with the evolutionary theory of a 24h day cycle for humans.

19.1.2 28(b) Sketch the distributions involved

```
from scipy.stats import norm
import numpy as np
import matplotlib.pyplot as plt
```

1. Finding the cutoff (try and error)

   ```
   print(1 - norm.cdf(26.35, 24, 1.2), norm.cdf(21.65, 24, 1.2))
   ```

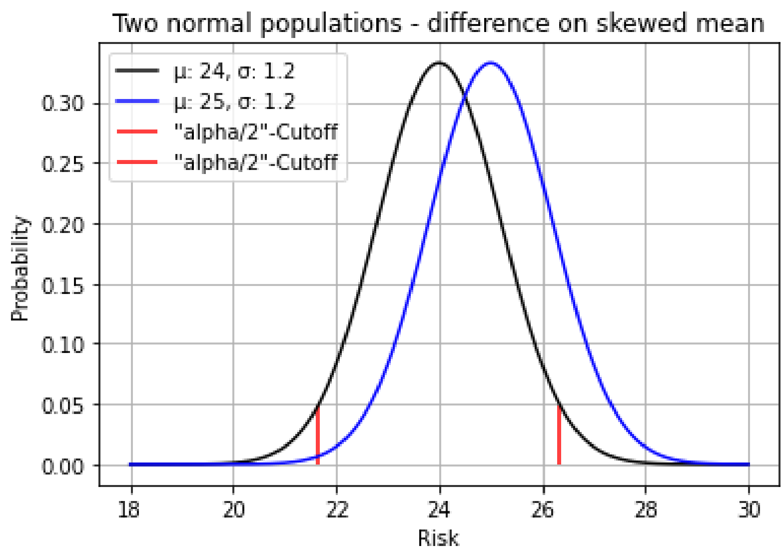
2. Sketch two distributions

   ```
   #x-axis
   x = np.arange(18, 30, 0.001)

   #define normal values (not normalized)
   plt.plot(x, norm.pdf(x, 24, 1.2), label=': 24, : 1.2', color='k')
   plt.plot(x, norm.pdf(x, 25, 1.2), label=': 25, : 1.2', color='blue')
   plt.vlines(x = 21.65, ymin=0, ymax=norm.pdf(21.65, 24, 1.2), colors = 'red', label
   plt.vlines(x = 26.35, ymin=0, ymax=norm.pdf(26.35, 24, 1.2), colors = 'red', label

   # # Grid on
   plt.grid(True)
   # Title
   plt.title('Two normal populations - difference on skewed mean')
   # Axis titles
   plt.xlabel('Risk')
   ```

```
plt.ylabel('Probability')
#add legend to plot
plt.legend()
```



19.1.3 28(c) Explain your answer to someone who has never taken a course in statistics

The results found support the theory. *A priori* to the test, we choose that we would only accept an alternative explanation if the tests showed that the results are so different from the expected that if we were to repeat them one hundred times and only would get five as extreme results at random or less.

In the research, we found that we would get nine false positives out of one hundred, at random. So, the results are not convincing enough to disprove the theory.

20 Exercise 29

"Five people who were convicted of speeding were ordered by the court to attend a workshop. A special device put into their cars kept records of their speeds for 2 weeks before and after the workshop. The maximum speeds for each person during the 2 weeks after the workshop follow.

Participant	Before	After
L.B.	65	58
J.K.	62	65
R.C.	60	56
R.T.	70	66
J.M.	68	60

Using the 5% significance level, should we conclude that people are likely to drive more slowly after such a workshop? (a) Use the steps of hypothesis testing. (b) Sketch the distributions involved. (c) Explain your answer to someone who is familiar with hypothesis testing involving known populations, but has never learned anything about t-tests."

20.1 Solution

20.1.1 29(a) Hypothesis testing steps

- Step 1: Restate the Question as a Research Hypothesis.
- Step 2: Determine the Characteristics of the Comparison Distribution.
- Step 3: Determine the Cutoff Sample Score on the Comparison Distribution at Which the Null Hypothesis Should Be Rejected.
- Step 4: Determine Your Sample's Score on the Comparison Distribution.
- Step 5: Decide Whether to Reject the Null Hypothesis.

1. S1: Directional P1: Before the workshop P2: After the workshop

$$\begin{cases} H_0 & : \overline{X_{\text{After}}} - \overline{X_{\text{Before}}} = 0 \\ H_1 & : \overline{X_{\text{After}}} - \overline{X_{\text{Before}}} < 0 \end{cases} \tag{6}$$

2. S2: p-values and paired t-statistic

   (a) Theoretical discussion We will be using a Paired T-Test. Because there is few measures of a sample, and we have a same population being tested in two different points in time.

   So, we have:

   - $\overline{X_{\text{After}}}$
   - $\overline{X_{\text{Before}}}$

- $t = \dfrac{\Delta\overline{X}}{S_D/\sqrt{n}}$
- $n = 5$

We will use the already implemented libraries, so there is no mistake.

(b) Numerical libraries Using the numerical library and the scientific libraries in Python,

```
import numpy as np
import scipy.stats as st

# State the data
data_before =[65,62,60,70,68]
data_after = [58,65,56,66,60]

tStat, pValue =  st.ttest_rel(data_before,data_after)
print("P-Value:{0:.2f} T-Statistic:{1:.2f}".format(pValue,tStat))
```

RESULTS: P-Value: 0.11 T-Statistic: 2.08

3. S3: Cutoff We will test mean differences for the same population. Also, we'll use $\alpha = 0.05$. As it's directional.

   $t_{(0.95,4)} = 2.132$. If the absolute value of the test statistic is greater than the critical value (0.95), then we reject the null hypothesis.

4. S4: Comparison to our data We had that the p-value = 0.11. As $p_{\text{value}} > \alpha$ (directional) we fail to reject H_0. Also, we could compare t_{obtained} and compare to see that as $t_{\text{obtained}} = 2.08 < t_{(0.95,4)} = 2.132$ then we fail to reject the null hypothesis.

5. S5: Conclusion So, the means of the populations are, in fact, the same, under our test and scrutiny. There is evidence that the mean velocities do not observe a decrease after the subjects go through the workshop.

20.1.2 29(b) Sketch the distributions involved.

1. Work with the data

```
from scipy.stats import norm
import numpy as np
import matplotlib.pyplot as plt
```

```
data_before =[65,62,60,70,68]
data_after = [58,65,56,66,60]

mean_before = np.mean(data_before)
mean_after = np.mean(data_after)
std_before = np.std(data_before)
std_after = np.std(data_after)
```

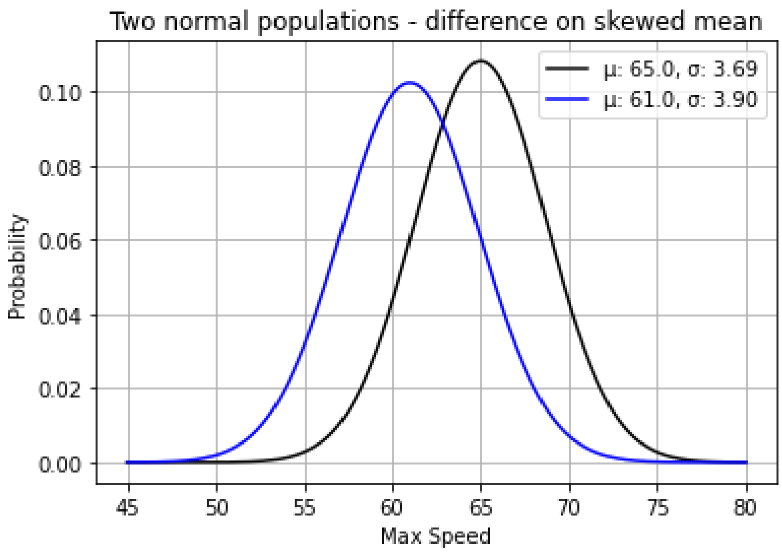
2. Sketch the two distributions

```
#x-axis
x = np.arange(45, 80, 0.001)

# 65.0 3.687817782917155 61.0 3.8987177379235853

#define normal values (not normalized)
plt.plot(x, norm.pdf(x, mean_before, std_before), label=': 65.0, : 3.69', color='k
plt.plot(x, norm.pdf(x, mean_after, std_after), label=': 61.0, : 3.90', color='blu
# plt.vlines(x = 21.65, ymin=0, ymax=norm.pdf(21.65, 24, 1.2), colors = 'red', lab
# plt.vlines(x = 26.35, ymin=0, ymax=norm.pdf(26.35, 24, 1.2), colors = 'red', lab

# # Grid on
plt.grid(True)
# Title
plt.title('Two normal populations - difference on skewed mean')
# Axis titles
plt.xlabel('Max Speed')
plt.ylabel('Probability')
#add legend to plot
plt.legend()
```



20.1.3 29(c) Explain your answer to someone who knows hypothesis testing

The results found support that there is no measurable difference before and after the educational workshop. The p-value was 0.11, when we *a priori* have set it to p-critical = 0.05.

This p-value is derived from a statistic pretty similar to the Z-statistic, but with estimated mean and variance. Also, the fact that the population is the same is also considered into the statistic.

Therefore, the results are not convincing enough to prove the workshop is, in fact, effecient.

21 Exercise 30

"For each of the following studies, say whether you would use a t-test for dependent means or a t-test for independent means. (a) A researcher randomly assigns a group of 25 unemployed workers to receive a new job-skills program and 24 other workers to receive the standard job-skills program, and then measures how well they all do on a job-skills test. (b) A researcher measures self-esteem in 21 students before and after taking a difficult exam. (c) A researcher tests reaction time of each of a group of 14 individuals twice, once while in a very hot room and once in a normal-temperature room."

21.1 Solution

21.1.1 30 (a)

A independent, paired, t-test would be proper. The populations are choosen at random and have no relation to each other

21.1.2 30 (b)

A dependent t-test would be proper. Because we are measuring the same variables, whitin the same population, in two distinctic points in time.

21.1.3 30 (c)

A dependent t-test because the same population is being compared in two different circunstance.

22 Exercise 31

"Figure SDifference for each of the following studies:

	N1	S²1	N2	S²2
a.	30	5	20	4
b.	30	5	30	4
c.	30	5	50	4
d.	20	5	30	4
e.	30	5	20	2

22.1 Solution

We are given $(N_1,\, S_1^2,\, N_2,\, S_2^2)_i$ for $i \in \{a, b, c, d, e\}$.

We know that

$$\begin{cases} df_{\text{Total}} = \sum_{i=0}^{n} df_i \\ df_i = N_i - 1 \\ S^2_{Pooled} & = \dfrac{df_1}{df_{\text{Total}}}\left(S_1^2\right) + \dfrac{df_2}{df_{\text{Total}}}\left(S_2^2\right) \\ S^2_{M_1} & = \dfrac{S^2_{Pooled}}{N_1} \\ S^2_{M_2} & = \dfrac{S^2_{Pooled}}{N_2} \\ S^2_{\text{Difference}} & = S^2_{M1} + S^2_{M2} \Leftrightarrow S_{\text{Difference}} = \sqrt{S^2_{M1} + S^2_{M2}} \end{cases} \tag{7}$$

22.1.1 Python program to automate the problem

We will create a python function to solve the problem

```
def calc_Sdiff(N1,S1_sqd,N2,S2_sqd):
    Spooled_sqd=((N1-1)/(N1+N2-2))*S1_sqd + ((N2-1)/(N1+N2-2))*S2_sqd
    Sm1_sqd=Spooled_sqd/N1
    Sm2_sqd=Spooled_sqd/N2
    Sdiff=np.sqrt(Sm1_sqd+Sm2_sqd)
    print(Sdiff)
```

22.1.2 31(a)

```
calc_Sdiff(30,5,20,4)
```

RESULTS: 0.62

22.1.3 31(b)

```
calc_Sdiff(30,5,30,4)
```

RESULTS: 0.55

22.1.4 31(c)

```
calc_Sdiff(30,5,50,4)
```

RESULTS: 0.48

22.1.5 31(d)

```
calc_Sdiff(20,5,30,4)
```

RESULTS: 0.61

22.1.6 31(e)

```
calc_Sdiff(30,5,20,2)
```

RESULTS: 0.56

23 Exercise 32

"For each of the following experiments, decide if the difference between conditions is statistically significant at the .05 level (two-tailed).

	Experimental Group			Control Group		
	N	M	s^2	N	M	s^2
(a)	10	604	60	10	607	50
(b)	40	604	60	40	607	50
(c)	10	604	20	40	607	16

23.1 Solution

The general formula for t is:

$$t = \frac{M_1 - M_2}{\sqrt{\frac{(N_1 - 1)(S_1^2) + (N_2 - 1)(S_2^2)}{N_1 + N_2 - 2}\left(\frac{1}{N_1} + \frac{1}{N_2}\right)}} \tag{8}$$

Let's create a Python algorithm for this.

We will also take $|t|$ and consider the $t_{\nu,0.975} = -t_{\nu,0.025}$ (two-sided).

```
def t_calc(N1,M1,S1_sqrd,N2,M2,S2_sqrd):
    df_total=N1+N2-2
    std1 = ((N1-1)*S1_sqrd)/df_total
    std2 = ((N2-1)*S2_sqrd)/df_total
    rev_mean = (1/N1 + 1/N2)
    t=(M1-M2)/np.sqrt((std1+std2)*rev_mean)

    print(abs(t))
```

23.1.1 32 (a)

```
t_calc(10,604,60,10,607,50)
```

```
0.9045340337332909
```

RESULTS: 0.90

Taking the value of $t_{18,0.975} = 2.101$ on a table.

As $t_{observed} < t_{18,0.975}$, It's not statistically significant.

23.1.2 32 (b)

```
t_calc(40,604,60,40,607,50)
```

RESULTS: 1.81

Taking the value of $t_{78,0.975} \approx t_{80,0.975} = 1.990$ on a table.

As $t_{observed} < t_{78,0.975}$, It's not statistically significant. Yet closer than last test.

23.1.3 32(c)

```
t_calc(10,604,20,40,607,16)
```

RESULTS: 2.073

Taking the values of $t_{48,0.975} \approx \frac{(t_{40,0.975} + t_{60,0.975})}{2} = \frac{2.021 + 2.000}{2} =$ 2.011 on a table.

As $t_{observed} > t_{48,0.975}$, It's statistically significant.

24 Exercise 33

"Twenty students randomly assigned to an experimental group receive an instructional program; 30 in a control group do not. After 6 months, both groups are tested on their knowledge. The experimental group has a mean of 38 on the test (with an estimated population standard of 3); the control group has a mean of 35 (with an estimated standard deviation of 5). Using the .05 level, what should the experimenter conclude? (a) Use the steps of hypothesis testing, (b) explain your answer to someone who is familiar with the t test for a single sample, but not with the t-test for independent means."

24.1 Solution

24.1.1 33(a) - Hypothesis steps

- Step 1: Restate the Question as a Research Hypothesis.
- Step 2: Determine the Characteristics of the Comparison Distribution.
- Step 3: Determine the Cutoff Sample Score on the Comparison Distribution at Which the Null Hypothesis Should Be Rejected.
- Step 4: Determine Your Sample's Score on the Comparison Distribution.

- Step 5: Decide Whether to Reject the Null Hypothesis.

1. S1: Is there differences in means? (undirectional) Population 1: Students who received instructional programs; Population 2: Students who haven't received instructional programs;

$$\begin{cases} H_0 & : \Delta\hat{\mu} = 0 \\ H_1 & : \Delta\hat{\mu} \neq 0 \end{cases} \tag{9}$$

2. S2: Data characteristics We have the following estimated statistical values,

$$(\hat{\mu_1} = 38,\ \hat{\sigma_1} = 3,\ \hat{\mu_2} = 35,\ \hat{\sigma_2} = 5) \tag{10}$$

3. S3: Cutoff The alpha level, $\alpha = 0.05$

   And the critical t-value:

$$t_{48,0.975} \approx \frac{(t_{40,0.975} + t_{60,0.975})}{2} = \frac{2.021 + 2.000}{2} = 2.011 \tag{11}$$

4. S4: Apply the t-test to our data, Applying the t-test of two independent means, and using the last's problems program written in Python,

```
t_calc(20,38,3,30,35,5)
```

   RESULTS: $t_{observed} = 5.066$

5. S5: Decide to reject the null hypothesis Because $t_{observed} = 5.066 \gg t_{48,0.975} = 2.011$, we can say with great confidence that there is a difference between the performance of people who took the instructional program.

24.1.2 33(b) - Explain the test for someone who knows t-test for a single sample

Just as we estimate the mean and variances of the sample, when we *a priori* choose a given value to test out the one sample t-test, we also use estimated values but we go a step further an give a experimental estimated value *a posteriori* to the control group.

Thus, following the modus operandi of comparing the $t_{observed}$ to a given table, we found that the alternative hypothesis was strongly significant. Taking the instructional program changes notoriously the outcomes.

25 Exercise 34

"What are the approximate numbers of participants needed for each of the following planned studies to have 80% power, assuming equal numbers in the two groups and all using the .05 significance level? (Be sure to give the total number of participants needed, not just the number needed for each group.)

Expected	-	-	-	-
	µ1	µ2		Tails
a.	10	15	25	1
b.	10	30	25	1
c.	10	30	40	1
d.	10	15	25	2

25.1 Solution

Using python's numerical library

```
from statsmodels.stats.power import TTestIndPower
```

25.1.1 34(a)

```
# standard deviation
std=25
# means of the samples
u1, u2 = 10, 15

# calculate the effect size
d = (u1 - u2) / std
print(f'Effect size: {d}')

# factors for power analysis
alpha = 0.05
power = 0.8

power = tt.TTestPower()
n_test = power.solve_power(nobs=None, effect_size = d,
   power = 0.8, alpha = 0.05, alternative='smaller')
n=2*n_test
print('Total number: {:.3f}'.format(n))
```

```
Effect size: -0.2
Total number: 311.851
```

RESULTS:

- Sample size: 312

25.1.2 34(b)

```
# standard deviation
std=25
# means of the samples
u1, u2 = 10, 30

# calculate the effect size
d = (u1 - u2) / std
print(f'Effect size: {d}')

power = tt.TTestPower()
n_test = power.solve_power(nobs=None, effect_size = d,
   power = 0.8, alpha = 0.05, alternative='smaller')
n=2*n_test
print('Total number: {:.3f}'.format(n))
```

RESULTS:

- Sample size: 23

25.1.3 34(c)

```
# standard deviation
std=40
# means of the samples
u1, u2 = 10, 30

# calculate the effect size
d = (u1 - u2) / std
print(f'Effect size: {d}')

power = tt.TTestPower()
n_test = power.solve_power(nobs=None, effect_size = d,
   power = 0.8, alpha = 0.05, alternative='smaller')
```

```
n=2*n_test
print('Total number: {:.3f}'.format(n))
```

RESULTS:

- Sample size: 53

25.1.4 34(d)

```
# standard deviation
std=25
# means of the samples
u1, u2 = 10, 15

# calculate the effect size
d = (u1 - u2) / std
print(f'Effect size: {d}')

power = tt.TTestPower()
n_test = power.solve_power(nobs=None, effect_size = d,
   power = 0.8, alpha = 0.05)
n=2*n_test
print('Total number: {:.3f}'.format(n))
```

RESULTS:

- Sample size: 397

26 Exercise 35

"An organizational psychologist was interested in whether individuals working in different sectors of a company differed in their attitudes towards the company. The results for the three people surveyed in engineering were 10, 12, and 11; for the three in the marketing department, 6, 6, and 8; for the three in accounting, 7, 4, and 4; and for the three in production, 14, 16, and 13 (higher numbers mean more positive attitudes). Was there a significant difference in attitude toward the company among employees working in different sectors of the company at the .05 level? (a) Use the steps of hypothesis testing. (b) explain your answer to someone who understands everything involved in conducting a t test for independent means, but is unfamiliar with the analysis of variance."

26.1 Solution

26.1.1 35 (a) Hypothesis Steps

- Step 1: Restate the Question as a Research Hypothesis.
- Step 2: Determine the Characteristics of the Comparison Distribution.
- Step 3: Determine the Cutoff Sample Score on the Comparison Distribution at Which the Null Hypothesis Should Be Rejected.
- Step 4: Determine Your Sample's Score on the Comparison Distribution.
- Step 5: Decide Whether to Reject the Null Hypothesis.

1. S1: Hypothesis undirectional ANOVA Let population 1, 2, 3 and 4 be the populations of Engineers, Marketing, Accountants and Production's departments.

   The Null and alternative hypothesis follow,

$$\begin{cases} H_0 : \hat{\mu_1} & = \hat{\mu_2} = \hat{\mu_3} = \hat{\mu_4} \\ H_1 : \hat{\mu_i} & \neq \hat{\mu_j}, \quad i \neq j \end{cases} \tag{12}$$

2. S2: Compare multiple instances of means; F-degrees We will use ANOVA - Analysis of Variance - due to the fact that are more than two measures to be tested against each other. The degrees of freedom for the F-distribution are:

$$\begin{cases} df_{between} & = 3 \\ df_{within} & = 8 \end{cases} \tag{13}$$

3. S3: Cutoff Looking at a F-table, $F_{(3,8),\,\alpha=0.05} = 8.8452$.
4. S4: Calculate out F-observed Using Python, the powerful scipy.stats have a Oneway F test ready,

```
from scipy.stats import f_oneway
```

   State the data,

```
engineers=[10,12,11]
marketing=[6,6,8]
accountant=[7,4,4]
production=[14,16,13]
```

Solve the problem,

```
f_oneway(engineers,marketing,accountant,production)
```

RESULTS: $\mathrm{F}_{\mathrm{onewayResult}}$(statistic=27.98550724637679, pvalue=0.00013597313862900978)

5. S5: Reject Null Hypothesis $T_{observed} = 27.99 \gg F_{(3,8),\,\alpha=0.05} = 8.8452$ therefore, with great confidence, we can say there is a difference in attitude in a company, dependent on sector.

26.1.2 35 (b) Explain to who know t-tests

A way to systematically tests multiple populations with equivalent t-tests two-by-two would be to use the ANOVA.

In this test, we can find a cutoff by looking at the degrees of freedom in between groups (4 different one, leading to 3 degrees of freedom) and within groups (3 different mesures for 4 different groups giving a total of 8 degrees of freedom); finally the $\alpha = 0.05$.

We test the variances of variance among and within groups, and make a ratio of them. As the result should that the variance due to variance among groups is way greater than within groups, then we can assert that the effect of variability is due to these groups having different behaviour in general (means and/or variation).

27 Exercise 36

"Rosalie Friend (2001), an educational psychologist, compared three methods of teaching writing. Students were randomly assigned to three different experimental conditions involving different methods of writing a summary. At the end of the two days of instructions, participants wrote a summary. One of the ways it was scored was the percentage of specific details of information it included from the original material. Here is a selection from her article describing one of the findings: The effect of summarization method on inclusion of important information was significant: F(2, 144) = 4.1032, p < .019. The mean scores (with standard deviations in parentheses) were as follows: Argument Repetition, 59.6% (17.9); Generalization, 59.8% (15.2); and Self-Reflection, 50.2% (18.0). (p. 14.) Explain these results to a person who has never had a course in statistics. Also, using the information in the above description."

27.1 Solution

This means that the chance of the high variation of scores be as extreme as the found in the study by change is of 1,9%. That is, we would need to repeat the study 100 times to find 2 of them having such extreme values by chance (not due to the method efficacy).

The $F(2,144) = 4.1032$ can be used, together with the data of means and standard deviations of each groups to derive an indirect measure of how much the method contributed to the differences in scores. If this value is greater than $F(2,144)$, then method is efficient in producing different scores.

27.1.1 Calculus of F observed

$$\begin{aligned} F &= \frac{S_{Between}}{S_{Within}} \\ S_{Between} &= \sqrt{n.S_M^2} \\ S_M^2 &= \frac{\sum (M - GM)^2}{df_{between}} \\ S_{Within} &= \frac{\sum_{i=1}^{N} S_i^2}{N} \end{aligned} \tag{14}$$

In which n: number of individuals per group; N: number of groups. 3.(n-1)=144 => n=(144+3)/3=49; N-1=2 => N=3.

1. $S_{between}$

$$\begin{aligned} GM &= \frac{(59.6 + 39.8 + 50.2)}{3} = 49.9 \\ \implies S_M^2 &= \frac{(59.6 - 49.9)^2 + (39.8 - 49.9)^2 + (50.2 - 49.9)^2}{(3-1)} = 98.1 \\ \implies S_{Between} &= \sqrt{(\frac{144+3}{3}) \times 98.1} = 69.33 \end{aligned} \tag{15}$$

2. S_{Within} $S_{Within} = \dfrac{(17.9 + 15.2 + 18.0)}{3} = 17.0$

3. $F_{observed}$ $F_{observed} = \frac{S_{between}}{S_{within}} = \frac{69.33}{17.0} = 4.08$

   We conclude $F_{observed} < F_{(2,144),0.05}$. So, in theory this result is negative. To be significant the result should be above the table value.

28 Exercise 37

"A researcher wants to be sure that the sample in her study is not unrepresentative of the distribution of ethnic groups in her community. Her sample includes 300 whites, 80 African Americans, 100 Latinos, 40 Asians, and 80 others. In her community, according to census records, there are 48% whites, 12% African Americans, 18% Latinos, 9% Asians, and 13% others. Is her sample unrepresentative of the population in her community? (Use the .05 level)

- Carry out the steps of hypothesis testing.
- Explain these results to a person who has never had a course in statistics."

28.1 Solution

28.1.1 37(a) Hypothesis Steps

- Step 1: Restate the Question as a Research Hypothesis.
- Step 2: Determine the Characteristics of the Comparison Distribution.
- Step 3: Determine the Cutoff Sample Score on the Comparison Distribution at Which the Null Hypothesis Should Be Rejected.
- Step 4: Determine Your Sample's Score on the Comparison Distribution.
- Step 5: Decide Whether to Reject the Null Hypothesis.

1. S1: Hypothesis Restatement Population$_{\{1,2,3,4,5\}}$ are the White, African Americans, Latinos and Asian, and Others respectively.

   H_0 says that the samples are representative of the community distribution, regarding frequency.H_1 is that is does not.

2. S2: Characteristics The comparison is a Chi-squared, with degree of freedom, $df = 5 - 1 = 4$.

3. S3: Cutoff The standard is of $\alpha = 0.05$. Looking a table $\chi_{4,0.05} = 9.488$. So, we have to achieve a $\chi_{observed} > \chi_{4,0.95} = 9.488$ in order to negate H_0.

4. S4: Metrics on observed data Using Python's scipy.stats, we can easily derive these results

```
import scipy.stats as st
import numpy as np
```

Declaring the data, quantity per group and expected quantity per group,

```
qtt_per_gr = [300, 80, 100, 40, 80]
```

Generate the expected quantity per group $f_{expected_{\{i\}}} = f_i \times \sum n_i$ which n_i is the number of observed people in a community.

```
frq = [0.48,0.12,0.18,0.09,0.13]
expt_per_gr=[(frq[i] * np.sum(qtt_per_gr)) for i in range(len(frq))]

st.chisquare(qtt_per_gr,expt_per_gr)
```

RESULTS: Power$_{\text{divergenceResult}}$(statistic=5.662393162393162, pvalue=0.22581947016382237)

5. S5: Do not reject the Null Hypothesis The value observer for $\chi_{observed} = 5.662 < \chi_{4,0.95} = 9.488$. Therefore, the sample do properly represent the general population. 22 out of 100 time we do this experiment we would get as extreme differences of frequences in this population, by random.

28.1.2 37 (b)

The general population is well represented in this setup. It would not be rare to pick randomly people in the population and end up with the study's populations ratio.

29 Exercise 38

29.1 Solution

29.1.1 37(a) Hypothesis Steps

- Step 1: Restate the Question as a Research Hypothesis.

- Step 2: Determine the Characteristics of the Comparison Distribution.
- Step 3: Determine the Cutoff Sample Score on the Comparison Distribution at Which the Null Hypothesis Should Be Rejected.
- Step 4: Determine Your Sample's Score on the Comparison Distribution.
- Step 5: Decide Whether to Reject the Null Hypothesis.

1. S1: Hypothesis Restatement Population$_{\{1,2,3\}}$ are the populations who preference A, B and C, at first respectively.

   H_0 says that the popularity would be measure as equal regarding frequency.H_1 is that is does not happen, the "inate" preference remains, regardless of marketing.

2. S2: Characteristics The comparison is a Chi-squared, with degree of freedom, $df = 3 - 1 = 2$.

3. S3: Cutoff The standard is of $\alpha = 0.05$. Looking a table $\chi_{2,0.05} = 5.991$. So, we have to achieve a $\chi_{observed} > \chi_{2,0.95}$ in order to negate H_0.

4. S4: Metrics on observed data Using Python's scipy.stats,

   ```
   import scipy.stats as st
   import numpy as np
   ```

   Declaring the data, quantity per group and expected quantity per group,

   ```
   qtt_per_gr = [197,120,210]
   ```

   Generate the expected quantity per group $f_{expected_{\{i\}}} = f_i \times \sum n_i$ which n_i is the number of observed people in a community.

   ```
   frq = [1/3,1/3,1/3]
   expt_per_gr=[(frq[i] * np.sum(qtt_per_gr)) for i in range(len(frq))]

   st.chisquare(qtt_per_gr,expt_per_gr)
   ```

   RESULTS: Power$_{\text{divergenceResult}}$(statistic=26.94117647058824, pvalue=1.4118802443206298e-06)

5. S5: Reject the Null Hypothesis The value observer for $\chi_{observed} = 26.94 > \chi_{2,0.05} = 5.991$. Therefore, the expected frequency do poorly represent the general frequency observe. Therefore, it means the study hypothesis had failed. People didn't hold a preference, because of how subjects were presented.

29.1.2 38 (b)

This result shows that the only way that we would obtain these values of popularity distributions, if in fact presenting people before hand as equal had the causal role, is if we took a sample that only would occur 14 times out of 1000. That is, presenting people before hand do not determine popularity.

30 Exercise 39

"Below are results of a survey of a sample of people buying ballet tickets, laid out according to the type of seat they purchased and how regularly they attended. Is there a significant relation? (Use the .05 level.)

- Carry out the steps of hypothesis testing.
- Explain your answer to someone who has never had a course in statistics.

		Attendance	
		Regularly	Occasional
	Orchestra	20	80
Seating Category	Dress Circle	20	20
	Balcony	40	80

30.1 Solution

30.1.1 39(a) Hypothesis Steps

- Step 1: Restate the Question as a Research Hypothesis.
- Step 2: Determine the Characteristics of the Comparison Distribution.
- Step 3: Determine the Cutoff Sample Score on the Comparison Distribution at Which the Null Hypothesis Should Be Rejected.
- Step 4: Determine Your Sample's Score on the Comparison Distribution.

- Step 5: Decide Whether to Reject the Null Hypothesis.

1. S1: Hypothesis Restatement There are 3x2 design of variables, two nominal relating to frequency and three nominal related to seating category.

   H_0 says that the popularity would be measured as equal regarding frequency and seating category; that is, they are independent dimensions of behaviour. H_1 is that there exist in fact dependency between Seating Category and Frequency of attendence.

2. S2: Characteristics The comparison is a Chi-squared of independece, with degree of freedom,

$$\begin{aligned} df_1 = 3 - 1 = 2,\ df_2 = 2 - 1 = 1 \Leftrightarrow (df_1, df_2) &= (2, 1) \\ \implies df = df_1 \times df_2 = 2 * 1 = 2 \end{aligned} \tag{16}$$

3. S3: Cutoff The standard is of $\alpha = 0.05$. Looking a table $\chi_{2,0.05} = 5.991$. So, we have to achieve a $\chi_{observed} > \chi_{2,0.95}$ in order to negate H_0.

4. S4: Metrics on observed data Using Python's scipy.stats,

```
      import pandas as pd
import numpy as np
from scipy.stats import chi2_contingency

import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline
```

   Declaring the data format,

```
df = pd.DataFrame({
    'Attendence':np.concatenate((
np.array(['Regular']*20),
np.array(['Occasional']*80),
np.array(['Regular']*20),
np.array(['Occasional']*20),
np.array(['Regular']*40),
```

```
np.array(['Occasional']*80))),
    'SeatingCategory':np.concatenate((
np.array(['Orchestra']*100),
np.array(['DressCircle']*40),
np.array(['Balcony']*120)))})
df.head()
```

Seeing if the table was rightfully stated,

```
contigency= pd.crosstab(df['Attendence'], df['SeatingCategory'])
contigency
```

SeatingCategory Attendence	Balcony	DressCircle	Orchestra
Occasional	80	20	80
Regular	40	20	20

Generate the expected quantity per group $f_{expected_{\{i\}}} = \frac{n_i}{\sum n_j}$.

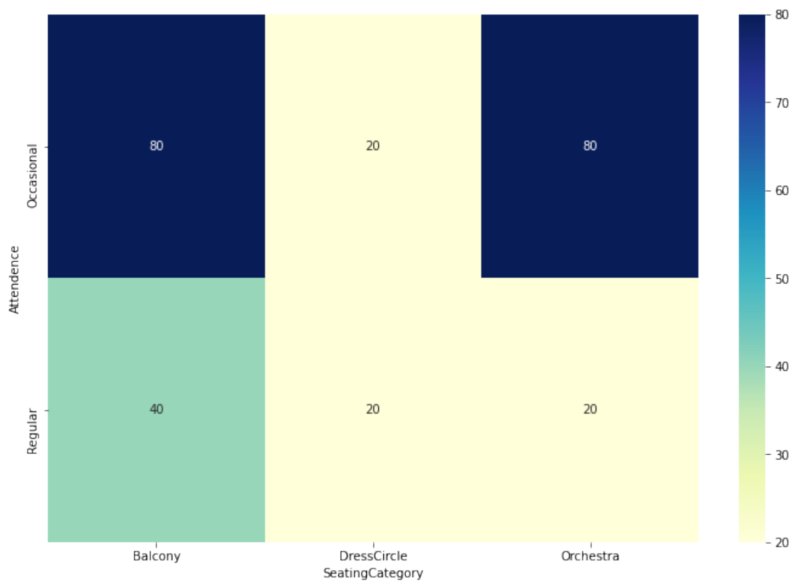
```
contigency_pct = pd.crosstab(df['Attendence'], df['SeatingCategory'], normalize='i
contigency_pct

# Chi-square test of independence.
c, p, dof, expected = chi2_contingency(contigency)
c,p
```

RESULTS: $(\chi, p_{value}) = (12.759259259259263, 0.0016957508962184467)$

Let's see the heatmap for this problem,

```
plt.figure(figsize=(12,8))
sns.heatmap(contigency, annot=True, cmap="YlGnBu")
```



5. S5: Reject the Null Hypothesis The value observer for $\chi_{observed} = 12.759 > \chi_{2,0.05} = 5.991$. Therefore, we can say that, in fact, there is an inbalance in the distribution of seats-type regarding frequency.

30.1.2 39 (b)

In other words, there is a relationship between frequency and the type of show one attend. Occasional attendents will prefer more Operas and Orchestras.

31 Exercise 40

"About how many participants do you need for 80% power in each of the following planned studies, using a chi-square test of independence with p < .05?"

	Predicted Effect Size	Design
(a)	Small	2x2
(b)	Medium	2x2
(c)	Large	2x2
(d)	Small	3x3
(e)	Medium	3x3
(f)	Large	3x3

31.1 Solution

We will be using statsmodels.stats.power.GofChisquarePower library in Python,

```
import statsmodels.stats.power as p
```

31.1.1 40(a) Small 2x2

```
p.GofChisquarePower().solve_power(effect_size=0.2,
alpha=0.05,
power=0.80,
n_bins=4)
```

RESULTS: 273

31.1.2 40(b) Medium 2x2

```
p.GofChisquarePower().solve_power(effect_size=0.5,
alpha=0.05,
power=0.80,
n_bins=4)
```

RESULTS: 44

31.1.3 40 (c) Large 2x2

```
p.GofChisquarePower().solve_power(effect_size=0.8,
alpha=0.05,
power=0.80,
n_bins=4)
```

RESULTS: 17

31.1.4 40(d) Small

```
p.GofChisquarePower().solve_power(effect_size=0.2,
alpha=0.05,
power=0.80,
n_bins=9)
```

RESULTS: 376

31.1.5 40(e) Medium

```
p.GofChisquarePower().solve_power(effect_size=0.5,
alpha=0.05,
power=0.80,
n_bins=9)
```

RESULTS: 60

31.1.6 40(f) Large

```
p.GofChisquarePower().solve_power(effect_size=0.8,
alpha=0.05,
power=0.80,
n_bins=9)
```

RESULTS: 24